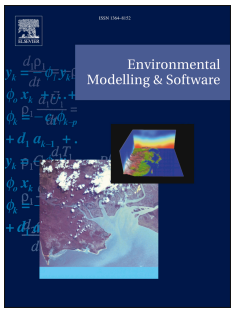
Accepted Manuscript



Modeling the extent of surface water floods in rural areas: Lessons learned from the application of various uncalibrated models

Daniel B. Bernet, Andreas Zischg, Volker Prasuhn, Rolf Weingartner

PII: S1364-8152(17)30857-5

DOI: 10.1016/j.envsoft.2018.08.005

Reference: ENSO 4270

To appear in: *Environmental Modelling and Software*

Received Date: 9 August 2017

Revised Date: 14 June 2018

Accepted Date: 3 August 2018

Please cite this article as: Bernet, D.B., Zischg, A., Prasuhn, V., Weingartner, R., Modeling the extent of surface water floods in rural areas: Lessons learned from the application of various uncalibrated models, *Environmental Modelling and Software* (2018), doi: 10.1016/j.envsoft.2018.08.005.

Modeling the extent of surface water floods in rural areas: lessons learned from the application of various uncalibrated models

Daniel B. Bernet[a,*], Andreas Zischg[a], Volker Prasuhn[b], Rolf Weingartner[a]

[a] *Institute of Geography & Oeschger Centre for Climate Change Research & Mobiliar Lab for Natural Risks, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland*
[b] *Agroscope, Research Division, Agroecology and Environment, Reckenholzstrasse 191, 8046 Zurich, Switzerland*

Abstract

Surface water floods (SWFs) do not only increasingly threaten cities, but also affect rural areas. So far, little research has been dedicated to the prediction of SWFs in rural environments, although in practice the process is already being considered in deterministic flood hazard assessments. To test the validity of such assessments, we select four raster-based models with differing complexity and evaluate whether they reliably predict inundated areas by SWF in rural areas. The uncalibrated models are first applied to four artificial surfaces and second, to eight case studies covering manifold geographical and meteorological settings. For the case studies, the models' prediction skills are assessed based on inundated areas inferred from various sources. The models' performance is rather low for all case studies, which highlights the necessity for calibration and/or validation of such models. Moreover, the case studies provide more general conclusions concerning the modeling of SWFs in rural areas.

Keywords: surface water flood, rural environment, flood inundation model, uncalibrated, validation

Software availability

- FLO-2D (cf. Sect. 2.1.1)
  - Details: FLO-2D Pro (Build No. 16.06.16)
  - Developers: FLO-2D Software Inc. (P.O. Box 66, Nutrioso, AZ 85932, United States)
  - Requirements: Windows 7 or higher
  - Cost: $995.00
  - URL: `https://www.flo-2d.com/flo-2d-pro/`
- FloodArea (cf. Sect. 2.1.2)

*Corresponding author
Email address: `daniel.bernet@giub.unibe.ch` (Daniel B. Bernet)

[1] Abbreviations: DEM, digital elevation model; GA, Green-Ampt; SWF, surface water flood; UAV, unmanned aerial vehicle

  - Details: FloodArea^{HPC}-Desktop 10.3 (4 Cores) on ArcGIS®10.3.0.4322
  - Developers: geomer GmbH (Im Breitspiel 11 b, 69126 Heidelberg, Germany)
  - Requirements: Windows 7 or higher and ArcGIS®10.3
  - Cost: €11'845.00 for 4 cores (€3'875.00 for 1 core)
  - URL: `http://www.geomer.de/en/software/floodarea`

- r.sim.water (cf. Sect. 2.1.3)
  - Details: Module r.sim.water in GRASS GIS 7.2.0 (2016)
  - Developers: H. Mitasova, J. Hofierka, C. Thaxton (and GRASS Development Team)
  - Requirements: Windows, Linux or Mac OSX
  - Cost: Free of charge (GNU General Public Licence)
  - URL: `https://grass.osgeo.org/grass72/manuals/r.sim.water.html`

- MFD (cf. Sect. 2.1.4)
  - Details: Tool Flow Accumulation (Top-Down) with option "Multiple Flow Direction" in SAGA GIS 4.1.0 (64-bit)
  - Developers: O. Conrad and T. Grabs
  - Requirements: Windows or Linux
  - Cost: Free of charge (GNU General Public Licence)
  - URL: `http://www.saga-gis.org/saga_tool_doc/4.1.0/ta_hydrology_0.html`

1. Introduction

Economic losses caused by floods have been heavily increasing over the past decades in absolute terms (Thieken et al., 2007; Kron et al., 2012; Grahn & Nyberg, 2017), mostly driven by societal development (e.g. Cutter & Emrich, 2005), but possibly exacerbated by climate change (Falconer et al., 2009; Barredo, 2009; Kundzewicz et al., 2014). In particular, the frequency and the intensity of heavy rainfall is expected to increase in many places in the future (Kundzewicz et al., 2014). This has drawn growing attention to surface water floods (SWFs), which are caused by intense rainfall that cannot be drained altogether by means of natural and/or artificial drainage systems, stem from surcharged sewers, channels, culverted watercourses or groundwater springs and, consequently, result in ponded water and overland flow (Hankin et al., 2008; Falconer et al., 2009). With a particular high percentage of impermeable areas, cities are particularly prone to SWFs, as exemplified by recent devastating flood events affecting urbanized areas in Western Europe,

such as Hull, UK (Pitt, 2008; Coulthard & Frostick, 2010), Copenhangen, DK (Haghighatafshar et al., 2014), Amsterdam, NL (Gaitan et al., 2016; Spekkers et al., 2017) or Münster, DE (Spekkers et al., 2017). Cities in developing countries are even more severely impacted, as examples from Southeast Asia (Chan et al., 2012; Hénonin et al., 2013) or Africa (Di Baldassarre et al., 2010; Kundzewicz et al., 2014) illustrate.

Not surprisingly, a lot of research is dedicated to urban areas in terms of modeling SWFs (e.g. Maksivomić et al., 2009; Chen et al., 2012; Sampson et al., 2013; de Almeida et al., 2016), flood loss estimation (e.g. Merz et al., 2010; Jongman et al., 2012; van Ootegem et al., 2015) as well as flood risk assessment and management (e.g. Kaźmierczak & Cavan, 2011; Blanc et al., 2012; Zhou et al., 2012; Löwe et al., 2017). In contrast, relatively little research has been dedicated to rural areas, in spite the fact that such areas are highly exposed to flooding, as examples from the European Alps point out (Fuchs et al., 2015, 2017; Röthlisberger et al., 2017). At the same time, these areas are not only affected by fluvial floods, but similarly by SWFs (Bernet et al., 2017). Moreover, overland flow generated on rural or peri-urban areas may be transferred into urbanized areas and thereby contribute to the adverse effects of SWFs within the urban environment, as well (Andrieu et al., 2004; Yu & Coulthard, 2015). Thus, scientific studies regarding the link between SWFs and assets at risk in rural areas are generally lacking.

In contrast, the topic of how to prepare for and manage SWFs has been discussed outside of academia (Bernet et al., 2017). This has led to a wide range of manuals and guidelines regarding SWF hazard assessment, risk management and awareness raising at the point scale (e.g. Egli, 2007; Rüttimann & Egli, 2010), as well as on communal and regional scales (e.g. Castro et al., 2008; DWA, 2013; LUBW, 2016; CEPRI, 2014). Therein, the focus lies generally on the built environment and, thus, the rural areas are considered, as well.

In practice, the tools used for SWF hazard assessments, consist mainly of single deterministic simulations with two-dimensional (2D) flood inundation models (cf. Meon et al., 2009; Tyrna & Hochschild, 2010; Kipfer et al., 2015; Tyrna et al., 2017). This circumstance has certainly been influenced by the heavily increasing availability of high-resolution digital elevation models (DEMs), driven by advancing data collection techniques (Wechsler, 2007; Fewtrell et al., 2008; Sampson et al., 2012; Chen et al., 2012; Neal et al., 2012; Dottori et al., 2013; de Almeida et al., 2016; Savage et al., 2016a). At the same time, the applicability of hydrodynamic flood inundation models to finer resolutions has been supported by increasing computational power (Fewtrell et al., 2008; Hunter et al., 2008; Dottori et al., 2013; Savage et al., 2016b). However, the exploitation of high-resolution DEMs is still limited by computational constraints (Chen et al., 2012; Sampson et al., 2012; de Almeida et al., 2016; Savage et al., 2016a,b; Tyrna et al., 2017). Thus, rather simple flood inundation models are applied in practice for SWF hazard assessments, as they usually encompass large areas.

In general, a compromise is inevitable when applying a model, balancing spatial resolution, model complexity and computational efficiency (Horritt & Bates, 2001; Cook & Merwade, 2009; Fewtrell et al., 2008,

2011; Sampson et al., 2012; Neal et al., 2012; Dottori et al., 2013; Savage et al., 2016a,b). At the same time, it is not obvious how the specific choice influences the models' performance. Moreover, recent studies have pointed out that the uncertainty associated with flood inundation models fed with high-resolution DEMs are more complex than previously thought (Abily et al., 2016; Savage et al., 2016b). Meanwhile, the models' extreme precisions may provoke overconfidence in their results, which could ultimately lead to wrong decisions in flood risk management (Dottori et al., 2013; Savage et al., 2016a).

Wrong decisions can usually be prevented by rigorously evaluating the applied models (e.g. Jakeman et al., 2006; Bennett et al., 2013). However, appropriate data are crucial for this task. Yet, there is an eminent lack of observational data (Blanc et al., 2012; Neal et al., 2012; Spekkers et al., 2014; Yu & Coulthard, 2015; Gaitan et al., 2016; Rözer et al., 2016), which impairs the applicability of such deterministic modeling approaches. Even more so for SWFs in rural areas, where the lack of observational data is particularly pronounced (Yu & Coulthard, 2015). In practice, however, the lack of observational data does not appear to prevent the use of single deterministic simulations for SWF hazard assessments. On the contrary, examples of overland flow predictions in urban, peri-urban and rural areas indicate that it rather leads to the renouncement of model calibration and/or validation (cf. Meon et al., 2009; Tyrna & Hochschild, 2010; Kipfer et al., 2015; Tyrna et al., 2017). Such approaches are employed to produce large-scale SWF hazard maps, as the examples of Kipfer et al. (2015) and Tyrna et al. (2017) indicate. In practice, these maps are then being used to identify potentially affected assets, such as buildings, infrastructure, agricultural fields, etc.

Thus, the question arises whether the deterministic tools reportedly used today in hazard assessments, e.g., for compiling SWF maps, are fit for their purpose of predicting areas exposed to SWF under various conditions. In particular, it is unclear how well such a modeling approach performs if such tools are not conditioned and/or evaluated due to a lack of observational data. Using this starting point, the main goal of this study is to evaluate whether uncalibrated and unvalidated 2D models can reliably predict the extent of SWFs in rural environments, on which basis potentially exposed assets can be identified, for instance. Based on this evaluation, we are able to draw conclusions about the suitability of this modeling approach for assessing the extent of SWFs, as well as for modeling SWFs in rural areas, in general.

For that matter, we directly explore the models' predictive skills by comparing their outputs (Teng et al., 2017). In the style of other studies benchmarking 2D models (e.g. Fewtrell et al., 2011; Neal et al., 2012; Néelz & Pender, 2013), we apply the models to artificial and real-world test cases. As a first exercise, we apply the selected models to four artificial surfaces, inspired by Zhou & Liu (2002). In this highly controlled and simplified environment, the models can easily be compared and inherent model characteristics are revealed. In a second exercise, the models are applied to real-world case studies. To mimic the commonly used approach in practical SWF hazard assessments, we apply similar uncalibrated 2D models with varying complexity. Thereafter, we quantitatively assess the models' predictive skills regarding the flooded area using common binary pattern performance measures (Bennett et al., 2013). As there are no data about

flow depths, flow velocities or flow dynamics available, we only compare the simulated with the observed SWF extents. For that matter, we have reconstructed inundated areas based on various observational data sources.

The seven study sites encompass various topographies, slopes, land use etc., while each of the eight case studies is associated with either relatively heavy or weak rainfall, respectively. Thus, for the purpose of this study, we relax the definition of SWFs and include not only events triggered by heavy rainfall, bot also events associated with weak rainfall. All events have in common that overland flow was produced, which led or could have led to damages to the built and unbuilt environment along the flow paths. As per definition, the inundations did not originate from overtopping watercourses, but are directly triggered by effective rainfall (cf. Bernet et al., 2017, for a discussion of related terms).

2. Materials and methods

2.1. Models

In this study, we test three raster-based, 2D hydrodynamic flood inundation models, i.e., FLO-2D, FloodArea and r.sim.water. The models have been selected such that different levels of model complexity are covered, following Neal et al. (2012). From the wealth of available 2D hydrodynamic models (cf. Teng et al., 2017), we chose FloodArea and r.sim.water since they have reportedly been used in the field of flood hazard assessment covering large areas including rural environments (cf. Kipfer et al., 2015; Tyrna et al., 2017). FLO-2D was selected as it is a "hydro-inundation model", using a term from Yu & Coulthard (2015) describing models that consider hydrological processes and overland flow routing, at the same time. Moreover, it has the most complex flow routing scheme among the selected models. Finally, the model selection was complemented by a flow accumulation algorithm, i.e., the multiple flow direction (MFD) algorithm introduced by Freeman (1991). Such flow algorithms have manifold applications due to their striking simplicity (cf. López-Vicente et al., 2014; Alder et al., 2015). An overview of the models' features is provided by Table 1.

2.1.1. FLO-2D

FLO-2D is a distributed, physically based flood inundation model (O'Brian, 2009). Among the selected models, it is the most sophisticated one, as it makes use of the full dynamic wave approximation (O'Brian, 2009). FLO-2D has various modules which can be switched on or off, if desired. It incorporates an infiltration module with various available methods, whereas the Green-Ampt (GA) method based on Green & Ampt (1911) is the most sophisticated one.

Table 1: Model feature comparison. "Yes" indicates features or modules that can be directly assessed by the respective model, "no" highlights unavailable features, while "NA" describes features that are not applicable, i.e., the rainfall-related features for the flow accumulation algorithm MFD.

Feature, modules	**FLO-2D**	**FloodArea**	**r.sim.water**	**MFD**
Flow depth	yes	yes	yes	no
Flow velocity	yes	yes	no	no
Flow barrier	yes	yes	no	no
Unsteady rain	yes	yes	no	NA
Interception	yes	no	no	NA
Infiltration	yes	no	no	NA

2.1.2. FloodArea

FloodArea is a simplified hydrodynamic flood inundation model that is fully integrated into a Geographic Information System (GIS), i.e., ArcGIS®by ESRI, with the main purpose of calculating areas affected by floods (geomer, 2016). The model cannot directly account for losses such as interception and infiltration. Thus, these losses have to considered by reducing the corresponding rainfall input (cf. Table 1 and Sect. 2.3.5).

2.1.3. r.sim.water

The hydrodynamic model r.sim.water simulates overland flow with a path sampling method, which is implemented as a module in the open source GIS software GRASS (Mitasova et al., 2004; Neteler et al., 2012). Similar to FloodArea, r.sim.water cannot directly account for losses such as interception and infiltration. Moreover, unsteady rainfall cannot be modeled (cf. Table 1 and Sect. 2.3.5).

2.1.4. MFD

MFD is a multiple flow direction algorithm that assesses the flow paths based solely on a digital elevation model (DEM) (Quinn et al., 1991). We use the algorithm implemented in the open source System for Automated Geoscientific Analyses (SAGA) (Conrad et al., 2015). Among the selected models, MFD is the simplest approach that does not route any water, but instead assesses each cell's relative catchment area. Consequently, the model does not predict any flow depths and velocities (cf. Table 1), but assesses a static characteristic of the topography, i.e., the distributed flow accumulation areas. Note that prior to applying MFD to real-world case studies, sinks and pits of the respective DEM were filled, as discussed by Wechsler (2007). For that matter, we used the algorithm by Planchon & Darboux (2002) with a value of 0.01° for the minimal slope.

2.2. Artificial surfaces

To assess the performance of flow routing algorithms, Zhou & Liu (2002) defined four different mathematical surfaces and compared the calculated specific catchment area with the theoretically true values. The application of such algorithms on smooth artificial surfaces reveals distinct patterns and characteristics reflecting the algorithm's differing mathematical formulations (cf. Zhou & Liu, 2002; Seibert & McGlynn, 2007; Pilesjö & Hasan, 2014). We adapt this approach to flood inundation modeling. Even without a theoretically true value, the adaptation of this approach to SWF modeling reveals inherent model characteristics that might not be apparent otherwise. Therefore, as a first exercise, we apply the selected models to four artificial surfaces, i.e., to a plane (Eq. 1), a concave (Eq. 2), a convex (Eq. 3) and to a combined concave/convex surface (Eq. 4). We compiled corresponding raster DEMs of 250-by-250 cells and a resolution of 2 m. The elevations of the plane are given by

$$z = ax + by + c \tag{1}$$

where $a \approx -0.051$, $b \approx 0.141$, $c = 0$ for a prescribed slope of $s = 15$ % and an aspect of $a = 160°$; $0 \leq x \leq 250$, $0 \leq y \leq 250$. The concave surface is defined as

$$\frac{x^2}{a} + \frac{y^2}{b} + \frac{z^2}{c} = 1 \quad z < 0 \tag{2}$$

where $a = 998$, $b = -748.5$, $c \approx 191.5$; $-250 \leq x \leq 0$ and $0 \leq y \leq 250$. The convex surface is given by

$$\frac{x^2}{a} + \frac{y^2}{b} + \frac{z^2}{c} = 1 \quad z > 0 \tag{3}$$

where $a = 998$, $b = -748.5$, $c \approx 191.5$; $0 \leq x \leq 250$, $-250 \leq y \leq 0$. Finally, the combined concave/convex surface is defined as

$$\frac{x^2}{a} + \frac{y^2}{b} = \frac{z}{c} \quad z < 0 \tag{4}$$

where $a = 998$, $b = -748.5$, $c \approx 191.5$; $-250 \leq x \leq 0$ and $-250 \leq y \leq 0$.

The artificial surfaces are further manipulated. Two rows of the corresponding DEMs are incised by a minimum of 0.3 m, in order to represents a 4 m wide street crossing the surfaces from West to East. This incision enables to test and visualize the influence of structures in the landscape that can have major effects on overland flow paths. The top views of the four artificial surfaces are shown in Fig. 1.

For the artificial surfaces, a rain event lasting one hour with an intensity of 50 mmh^{-1} was simulated. Infiltration and interception losses were not considered. A Manning's roughness coefficient of $n = 0.24$ $\mathrm{sm}^{-1/3}$ was chosen for all artificial surfaces, which corresponds to the value recommended for dense grass by McCuen (2016). A value of $n = 0.012$ $\mathrm{sm}^{-1/3}$ is chosen for the incised streets, which corresponds to the recommended value for asphalt (McCuen, 2016).

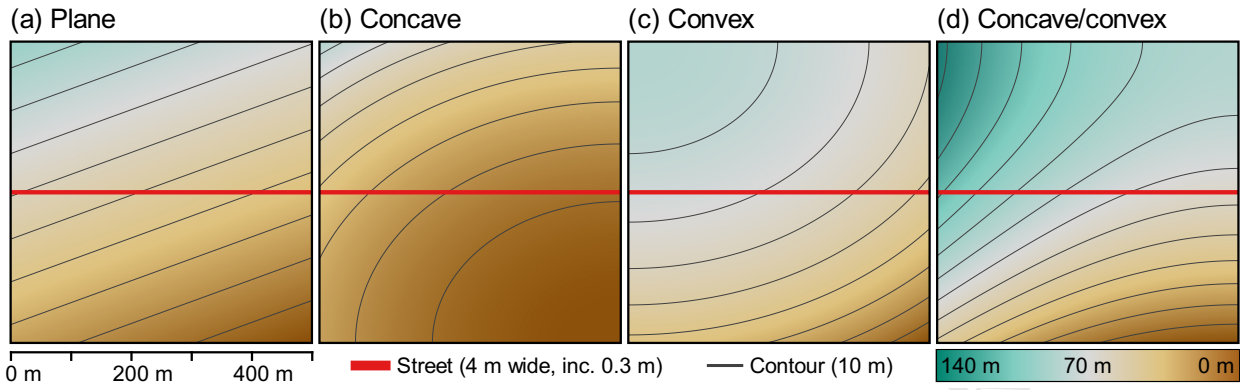


Figure 1: A plane (a, mean slope = 14.9 ± 1.6 %), concave (b, mean slope = 12.6 ± 6.5 %), convex (c, mean slope = 12.6 ± 6.5 %) and a combined concave/convex (d, mean slope = 20.5 ± 8.3 %) artificial surface used for an initial test of the models.

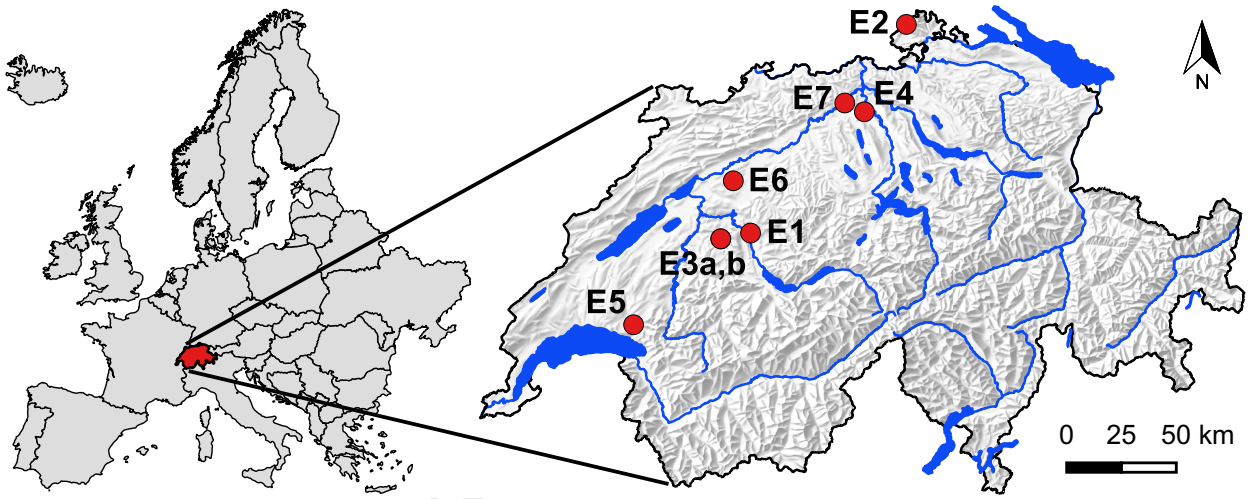


Figure 2: Location of the seven study sites. Note the two case studies (E3a and E3b) were observed at the corresponding study site.

2.3. Real-world case studies

We elaborated eight real-world case studies at seven study sites, i.e., at one study site two different events were observed. The case studies' characteristics are summarized in Table 2 and 3, while their respective location is shown in Fig. 2. In the following we introduce the delineation of the study perimeters, the gathered input data, the consideration of hydrological losses, the reconstruction of overland flow paths as well as the assessment of the model performance.

In terms of hydrological losses, we account for infiltration and interception losses, but neglect evaporation, as contributions of the latter are generally particularly low (cf. Yu & Coulthard, 2015). Furthermore, we assume that the influence of the sewer system on the flood extent is negligible. On the one hand, the fractions of built-up area are tiny in comparison to the rural areas for all case studies. On the other hand,

Table 2: Characteristics of the five different case studies (at four study sites) triggered by relatively heavy rainfall.

Characteristic	**E1**	**E2**	**E3a / E3b**	**E4**
Date	20.07.2007	02.05.2013	12.07.2014 / 06.06.2015	08.06.2016
Town and Canton (abbr.)	Rubigen BE	Schleitheim SH	Mittelhaeusern BE	Dottikon AG
Slope: mean ± sd (%)	7.7 ± 6.4	14.7 ± 9.3	18.1 ± 8.4	17.1 ± 10.1
Altitude: mean $\pm\Delta h/2$ (m)	571 ± 25	562 ± 62	727 ± 58	505 ± 82
Watershed domain D_{wsd} (km^2)	1.26	0.83	0.33	0.64
Observation domain D_{obs} (km^2)	0.61	0.22	0.33	0.33
Preconditions (-)	dry	normal	wet / dry	normal
Rainfall duration (h)	5	6	13 / 4	11
Rainfall sum (mm)	48.0	23.9	44.5 / 32.3	61.9
Max rainfall int. (mmh^{-1})	41.8	21.5	13.2 / 31.9	26.0
Mean rainfall int. (mmh^{-1})	9.6	4.0	3.4 / 8.1	5.6

Table 3: Characteristics of the three case studies triggered by relatively weak rainfall.

Characteristic	**E5**	**E6**	**E7**
Date	03.07.2007	13.05.2016	14.05.2016
Town and Canton (abbr.)	Bossonnens FR	Oberramsern SO	Oberflachs AG
Slope: mean ± sd (%)	12.7 ± 10.9	22.1 ± 20	22.9 ± 11.5
Altitude: mean $\pm\Delta h/2$ (m)	765 ± 42	586 ± 92	614 ± 110
Perimeter P_{wsd} (km^2)	0.28	0.25	0.54
Perimeter P_{obs} (km^2)	0.04	0.03	0.09
Preconditions (-)	wet	wet	normal
Rainfall duration (h)	49	27	68
Rainfall sum (mm)	56.5	59.7	90.1
Max rainfall int. (mmh^{-1})	6.7	7.3	6.5
Mean rainfall int. (mmh^{-1})	1.2	2.2	1.3

the field observations indicated that the sewer systems were often either blocked (e.g., by eroded material, branches, leaves, hail, etc.) or surcharged. Thus, in this study, the interactions between overland flow and the sewer systems are neglected, as assumed similarly by e.g. Fewtrell et al. (2011) or Kipfer et al. (2012).

2.3.1. Domains

In order to delineate the study perimeter for each case study, the area is considered, within which documented observations regarding overland flow paths are available. The corresponding study perimeters were obtained by delineating the smallest respective watershed that still encompassed the reconstructed flow paths. Thereafter, these perimeters were buffered by at least 50 m to obtain a simulation domain that extends over the watershed's boundary. This ensures that the simulations' boundary effects within the study perimeters remain negligible. Thus, three different domains are differentiated for each case study:

- Observation domain (D_{obs}), within which all documented overland flow paths were reconstructed.
- Watershed domain (D_{wsd}) representing the smallest watershed that contains the observation perimeter. The model results were cropped to this area.
- Simulation domain (D_{sim}) representing the buffered watershed domain, within which the simulations were carried out.

2.3.2. Primary input data

The main input for all four models is a DEM (Fig. 3). We used the DEM "swissALTI3D" as of 2013 with a regular grid size of 2-by-2 m, provided by the Swiss Federal Office of Topography (swisstopo, 2017a). Although, there are DEMs available with finer resolutions for some of the study sites, we used the aforementioned product, as it is homogeneous and available for whole Switzerland. As r.sim.water and MFD do not offer a direct option to integrate flow barriers such as buildings (cf. Table 1), the corresponding DEM was modified. All cells whose centroids were covered by a building were elevated by at least 10 m.

The land use was assessed between July 2014 and June 2016. As the land use was observed shortly after each event that falls into this period, i.e., E3a, E3b, E4, E6 and E7 (cf. Table 2 and 3), the corresponding land use represent the conditions during these events. In contrast, the land use of the remaining case studies were assessed roughly three years after the date of occurrence, or more. Although there is a slight time shift, we assumed that the mapped land use is representative for the respective case study, as major land use changes are not expected at these study sites within the respective period. Firstly, the land use including buildings, streets, fields, etc. was digitized using orthophotos from the product "SWISSIMAGE" (swisstopo, 2017b). Secondly, the land use was adjusted and verified based on field observations.

The surface roughness values were obtained by linking the land use with literature tables, i.e., with the comprehensive collection from McCuen (2016), as indicated in Fig. 3. The corresponding values are listed in Table 4.

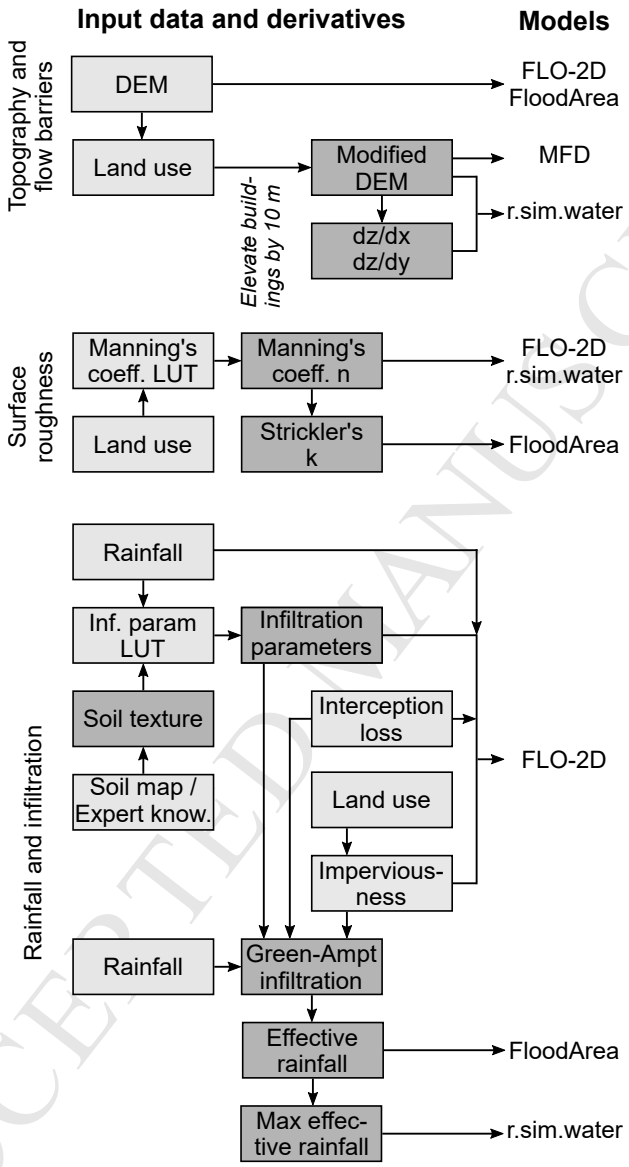


Figure 3: Primary model input data (light gray boxes) as well as derivatives thereof (dark gray boxes). The look-up table (LUT) for Manning's coefficients is based on values from McCuen (2016), whereas the Strickler values were obtained by taking their inverse, i.e., $k = 1/n$. The LUT for the infiltration parameters is based on Rawls et al. (1983) and O'Brian (2009). The Green-Ampt infiltration with ponding was implemented and calculated externally (cf. Sect. 2.3.3).

Table 4: Look-up table for relevant land use and corresponding Manning's roughness coefficients n, as recommended by McCuen (2016). The imperviousness is described by m (cf. Sect 2.3.5). It indicates, whether the corresponding cell is considered as being fully impervious ($m = 1$, no infiltration), partially impervious ($m < 1$, reduced infiltration) or completely pervious ($m = 0$, normal infiltration). Note that the rain falling on buildings did not contribute to the overland flow.

Land use	**Surface**	$\boldsymbol{n}$ ($\mathrm{sm}^{-1/3}$)	$\boldsymbol{m}$ (-)
Ley, meadow	Dense grass	0.240	0
Cropland	Conventional tillage without residue	0.090	0
Orchard	Woods without underbrush	0.200	0
Forest	Woods with light underbrush	0.400	0
Garden	Bermuda grass	0.410	0
Path, track	Graveled surface	0.012	0.75
Paved surface	Asphalt	0.012	1
Building	Smooth concrete	0.011	1

The hourly rainfall rate was extracted from the product "CombiPrecip" provided by the Federal Office of Meteorology and Climatology (MeteoSwiss, 2014). The product combines radar and rain gauge measurements by means of a co-kriging with external drift (e.g. Sideris et al., 2014; Panziera et al., 2016). It has a spatial resolution of 1-by-1 km, a temporal resolution of one hour and is available from 2005 onwards (MeteoSwiss, 2014). As the case study sites are small, each study perimeter is covered by just a few cells. To reduce the influence of single cells that might contain outliers, the raster cells covering each perimeter were buffered by one cell. Thereafter, the mean of these cells were calculated for each time step. Next, the triggering rainfall events were extracted from the rainfall records by considering a minimum inter-event time of $t_{min} = 6$ h and a minimum intensity threshold of $i_{min} = 0.1$ mmh^{-1}, which are in line with common literature values (e.g. Dunkerley, 2008). Consequently, at the beginning of each event, the rain intensity had been < 0.1 mmh^{-1} for at least six consecutive time steps of one hour each. Analogous, at the end of the event, it did not rain for at least six hours with an intensity ≥ 0.1 mmh^{-1}.

2.3.3. Infiltration

Out of all four models only FLO-2D allows the user to account for infiltration directly, while it cannot be modeled explicitly by FloodArea and r.sim.water, whereas MFD is not dependent on rainfall altogether (Table 1). Therefore, the following approach was chosen: the full potential of FLO-2D was exploited by using the integrated GA infiltration module. To feed the other two models with similar input, as recommended by

Neal et al. (2012), the GA method was implemented in R (R Core Team, 2016). Therewith, spatially and temporally variable cumulative infiltration rates were calculated. Based on these values, effective rainfall rates were obtained that were used as model inputs for FloodArea and r.sim.water (Sect. 2.3.5). Hereafter, the implementation and parametrization of the GA method are briefly outlined.

Based on Green & Ampt (1911), the cumulative infiltration $F(t)$ (mm) at time t (h) can be expressed as

$$F(t) = Kt + \Psi\Delta\Theta\left(\frac{F(t)}{\Psi\Delta\Theta} + 1\right) \quad , \tag{5}$$

whereas K is the hydraulic conductivity (mmh^{-1}), Ψ the wetting front soil suction head (mm), $\Delta\Theta = \Theta_f - \Theta_i$ (-) the difference between the final and initial soil moisture content. Thereby, an important assumption is that the water is ponded at the surface from the beginning of the steady rainfall. As this is generally not the case, Mein & Larson (1973) extended the GA infiltration method to account for the time until water starts to pond ($t = t_p$), at which time the cumulative infiltration depth equals the cumulative rainfall. Accordingly, the cumulative infiltration for steady rainfall after ponding time (i.e., $t > t_p$) is given by

$$F(t) = K(t - t_p) + F_p + \Psi\Delta\Theta\left(\frac{F(t)}{\Psi\Delta\Theta} + 1\right) \quad , \tag{6}$$

whereas t_p denotes the ponding time (h) and $F_p = F(t)$ the cumulative infiltration (mm) at ponding time $t = t_p$. We then implemented the GA method following Chu (1978), who expanded the method for unsteady rainfall events. The interested reader may refer to Chu (1978), who provides a detailed derivation and applied examples of the method.

The required GA infiltration parameters were obtained as follows: we estimated each study site's dominant soil texture based on expert knowledge, except for the case studies E4 and E6 for which soil maps including soil texture classes were available. We estimated the hydraulic conductivity K, the wetting front soil suction head Ψ and the effective porosity n_e using published regression parameter values from the comprehensive study by Rawls et al. (1983). Furthermore, it is assumed that the soils were saturated to a degree of $s_i = 30$, 50 or 80 % before each event under dry, normal or wet conditions, respectively. Each respective condition was set according to the observed antecedent rainfall (cf. Table 2 and 3). The change in soil moisture content was then estimated by $\Delta\Theta = n_e(s_f - s_i)$, while assuming that the soil's saturation after the event was $s_f = 100$ %.

2.3.4. Interception

Canopy storage capacity depends on various factors and roughly amounts 1 mm (e.g. Ward & Robinson, 2000). Thus, the depletion of this storage is tiny in comparison to the total rainfall volumes of the corresponding case studies (cf. Table 2 and 3). Moreover, as the values of different land cover types are within the same order of magnitude, we simply considered a bulk interception loss of 1 mm. In FLO-2D this loss volume

could be entered as a model parameter. For FloodArea and r.sim.water, we deducted the interception losses S (mm) from the total rainfall $P_t(t)$ (mm) to obtain a net rainfall $P_n(t)$ (mm) that reached the ground, as follows.

$$P_n(t) = \begin{cases} 0, & P_t(t) \leq S \\ P_t(t) - S, & P_t(t) > S \end{cases} \tag{7}$$

2.3.5. Effective rainfall

Infiltration cannot be modeled directly by FloodArea and r.sim.water (cf. Sect. 2.3.3). Thus, to account for infiltration and interception losses, we computed effective rainfall rates, which were then used as model inputs. The effective rainfall is given by

$$P_e(t) = P_n(t) - (1 - m)F(t) \quad , \tag{8}$$

whereas P_e is the cumulative effective rainfall (mm), P_n is the net rainfall that considers an initial interception loss (mm, cf. Eq. 7), m is the imperviousness factor (cf. Table 4) and $F(t)$ is the cumulative infiltration (mm, Sect. 2.3.3). Note that an imperviousness factor can be set directly in FLO-2D's GA infiltration module for each individual cell (O'Brian, 2009). However, for FloodArea and r.sim.water, the imperviousness factors as specified in Table 4 were considered during the assessment of cell- and time-specific effective rainfall rates.

In FloodArea, spatial variable rainfall can be modeled by providing weighting factors (geomer, 2016), which can be thought of as runoff coefficients relating the hyetograph to cell-specific effective rainfall. Obviously, these coefficients are changing over time and space. They are defined as $c_{i,j,t} = P_e(i,j,t)/P_t(t)$. The spatially variable rainfall can then be modeled by creating a raster with cell values $c_{i,j}(t)$ for each time step t. The simulations can then be stopped after each time step and restarted with the runoff coefficients of the next time step. This procedure was automated with batch scripts.

For r.sim.water, this procedure is not straightforward, as the simulations cannot be restarted based on results from a previous time step. Therefore, we chose the time step with the highest effective rainfall rate and ran the model with only this single spatially variable rainfall field.

2.3.6. Reconstruction of overland flow paths

Data sources that possibly indicate past SWFs include insurance claim records, disaster databases, reports and recollections from affected people (Bernet et al., 2017, and references therein). However, for recent events, it is usually possible to reconstruct flow paths of SWFs based on their traces in the field, as exemplified by Fig. 4. Particularly in rural environments, overland flow usually leaves notable traces such as erosion marks, deposited sediments and flattened vegetation. For the purpose of this study, we have reconstructed discernible SWF traces based on field observations following the events of the case studies

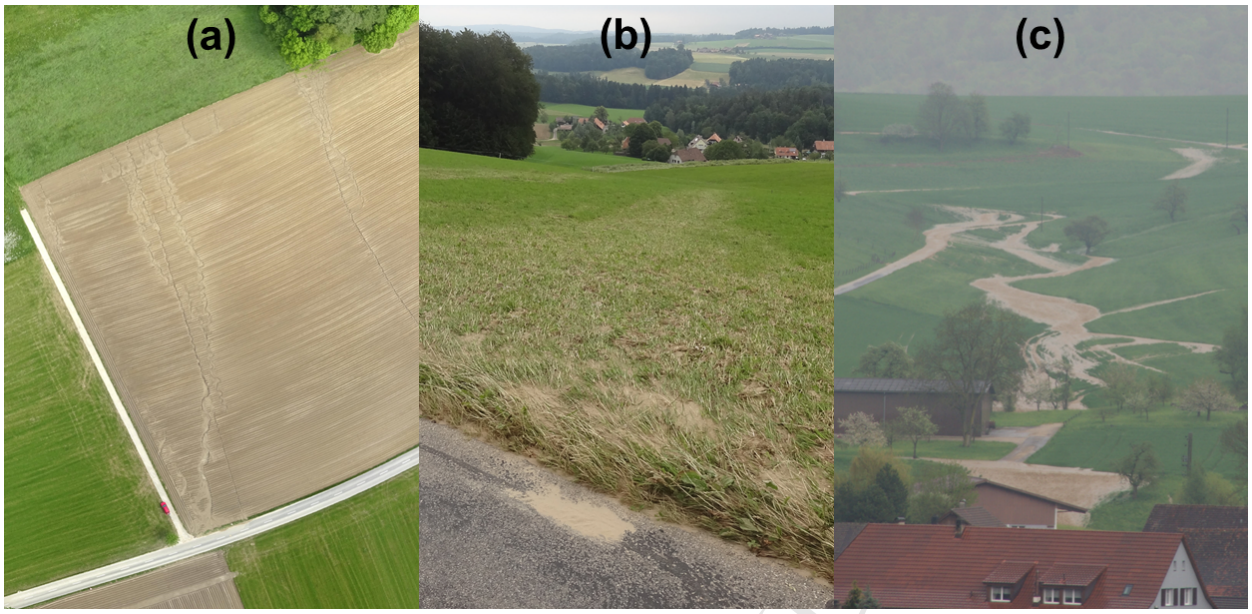


Figure 4: Different sources used for reconstructing overland flow paths. (a) Orthophoto derived by means of a unmanned aerial vehicle (UAV) documenting traces of erosion in a field of the case study E6 (source: Elias Hodel, 16.05.2016). (b) Traces of overland flow in a field of the case study E3b that were mapped in the field (source: Daniel B. Bernet, 09.06.2015). (c) Photograph documenting actual overland flow of the case study E2 (source: Andrea Wanner-Staubesand, 02.05.2013).

E3a, E3b, E4, E6 and E7, whereas for the remaining case studies, i.e., E1 and E2, the inundated areas were reconstructed based on external sources. Table 5 summarizes the source for the flow path reconstructions along with associated limitations, as well as a qualitative confidence level of the data quality.

Irrespective of the data source, the flow paths were reconstructed and spatially localized. Using standard GIS software, the field assessment were then digitally stored. All overland flow traces and paths were considered as being wet. To assess the performance of the models, these areas were compared to the model outputs, as outlined in the following section.

2.3.7. Model performance

Across various disciplines, map comparisons are a standard procedure to assess and compare model performances (e.g. Kuhnert et al., 2005; Foody, 2007; Bennett et al., 2013). However, there is not a single best method for this task. On the contrary, many tools including both quantitative as well as qualitative methods are recognized as being appropriate for this purpose (Kuhnert et al., 2005; Bennett et al., 2013). Thus, the model performance assessment have to be adapted to the models' objectives as well as to the characteristics of the available data, since the task is inherently case-specific (Bennett et al., 2013).

Along these lines, we compared the model outcomes and observations visually, as well as quantitatively. In terms of the latter, we used common binary pattern performance measures based on the contingency

Table 5: Exploited sources of information for reconstruction of inundated areas for each case study (cf. Tab. 2 and 3).

ID	Source	Quantity	Limitations	Confidence
E1	External map	Ponded water and water on the streets	No indication of flow paths, assessment methods unknown	Low
E2	Photographs	Flow paths and flood extent	Spatial localization of depicted flow paths	high
E3a,b	Field visits, aerial photos	Traces of flow (sediments, flattened vegetation)	Impossible to identify flow that left no traces	high
E4	Field visit	Traces of flow (sediments, flattened vegetation)	Impossible to identify flow that left no traces; Flow traces in forest difficult to detect	medium
E5	Video	Flow dynamics and extent of flood	Coverage limited to small area, low resolution	medium
E6	Field visit, aerial photos	Traces of erosion in bare field	Impossible to identify flow that left no traces	medium
E7	Field visit	Traces of erosion in bare field	Impossible to identify flow that left no traces; small observation perimeter compared to watershed	medium

Table 6: Contingency table of model prediction or observation (A) versus model prediction (B).

	Present (wet) in A	**Absent (dry) in A**
Present (wet) in B	Hits: $a = A_1B_1$	False alarms: $b = A_0B_1$
Absent (dry) in B	Misses: $c = A_1B_0$	Correct negatives $d = A_0B_0$

table (Table 6), which are widely being used for the comparison of simulated and observed flood extents (e.g. Aronica et al., 2002; Schumann et al., 2009; Stephens et al., 2014; Zischg et al., 2018). However, more recently, Stephens et al. (2014) pointed out that these performance measures are all subjected to a varying degree of bias, which should be considered in subsequent conclusions. As we are using different measures conjunctively in this study and are more interested in the broader picture, the influence of this circumstance on our conclusions is negligible.

The binary pattern performance measures are based on the assessment whether a cell was observed and/or simulated as wet or dry. All cells covered by an observed flow paths are considered as wet. For the simulation results, this information was inferred from the simulated maximum flow depths h_f (m) by applying an arbitrary threshold h_t (m). Thus, cells with a maximum flow depth below the threshold ($h_f < h_t$) are considered to be dry, while all other cells ($h_f \geq h_t$) are considered to be wet. We tested different threshold values and compared the performance of all models applied to all case studies using the observations as the reference. Based on these results, we empirically chose a value of $h_t = 0.02$ m as this threshold value maximized the performance of all models. Note that this threshold value is case-specific and, thus, might be different for other models, observational data, resolutions, etc.

In the following, we compare the models' results with observations, in addition to a comparison of the models among each other. The comparisons of the models with observations are constrained to the observation perimeter (D_{obs}), while the model comparison among each other is carried out within the whole watershed (D_{wsd}, cf. Sect. 2.3.1).

For the quantitative model comparison, we used the following binary pattern performance measures (e.g. Aronica et al., 2002; Pappenberger et al., 2007; Bennett et al., 2013), which are based on the contingency table (Table 6):

$$\text{Bias:} \quad m_1 = \frac{a+b}{a+c} \qquad m_1 \in [0,\infty] \qquad \text{ideally } m_1 = 1 \tag{9}$$

$$\text{Critical success index:} \quad m_2 = \frac{a}{a+b+c} \qquad m_1 \in [0,1] \qquad \text{ideally } m_2 = 1 \tag{10}$$

$$\text{Hit rate:} \quad m_3 = \frac{a}{a+c} \qquad m_1 \in [0,1] \qquad \text{ideally } m_3 = 1 \tag{11}$$

$$\text{False alarm rate:} \quad m_4 = \frac{b}{b+d} \qquad m_1 \in [0,1] \qquad \text{ideally } m_4 = 0 \tag{12}$$

Note that the critical success index (CSI) is also referred to as threat score or F^2 statistic in the literature (e.g. Bennett et al., 2013; Stephens et al., 2014).

3. Results

3.1. Artificial surfaces

Despite the lack of a baseline, applying the models to the selected artificial surfaces reveals interesting characteristics (Fig. 5). First and foremost, the incised street represents a prominent topographical structure that has a significant influence on the flow pattern. The street acts like a channel, which can collect incoming water and can be overtopped, if the channel is full or if the incoming water is not sufficiently deflected. Whether the street is overtopped or not, is discernible by the amount of dry cells directly to the south of the incised street, i.e., cells with a flow depth or flow accumulation below the flow threshold (dark red cells in Fig. 5). For each artificial surface, the pattern of these dry cells varies significantly among the models. In contrast, the pattern of dry cells north of the street is more similar among the models for all but the convex surface, as discussed later. Thus, the street has a major influence on the distribution of dry and wet cells, respectively.

In more detail, r.sim.water does not predict a deflection of the water crossing the street on any surface. Quite the opposite is true for the flow accumulation calculated by MFD. For all but the combined concave/convex surface, the street poses a complete or nearly complete flow barrier. FLO-2D and FloodArea, on the other hand, show a more differentiated picture, as water is overtopping where ever the flow depths are exceeding the street's incision. This is most apparent on the concave surface, where FloodArea predicts a significant overtopping of the street's eastern end, unlike the other models. Thus, in this modeling exercise, the user's choice of a model does not only heavily influence the pattern of dry and wet cells south of the street, but also the corresponding flow paths.

The results of the hydrodynamic models do not only deviate substantially south of the street, but also on the street itself for each artificial surface. FLO-2D consistently predicts the highest flow depths on the street. FloodArea's results exhibit flow depths that lie mostly between the minimal values estimated by r.sim.water and the high values predicted by FLO-2D. However, as mentioned before, a striking difference of FloodArea compared to FLO-2D is the overtopping of the street's incision at the eastern side of the concave surface. Compared to the other hydrodynamic models, r.sim.water predicts by far the lowest flow depths on the street for all surfaces. In fact, the flow depths on the street predicted by r.sim.water are below the wet/dry threshold of 0.02 m for all surfaces. Thus, the flow patterns south of the street are heavily influenced of how the models predict the flow over this topographical structure.

However, the flow patterns are also dependent on how the models simulate flow over the four different topographical forms. Specifically, the flow patterns on the convex surface of each single model is strikingly

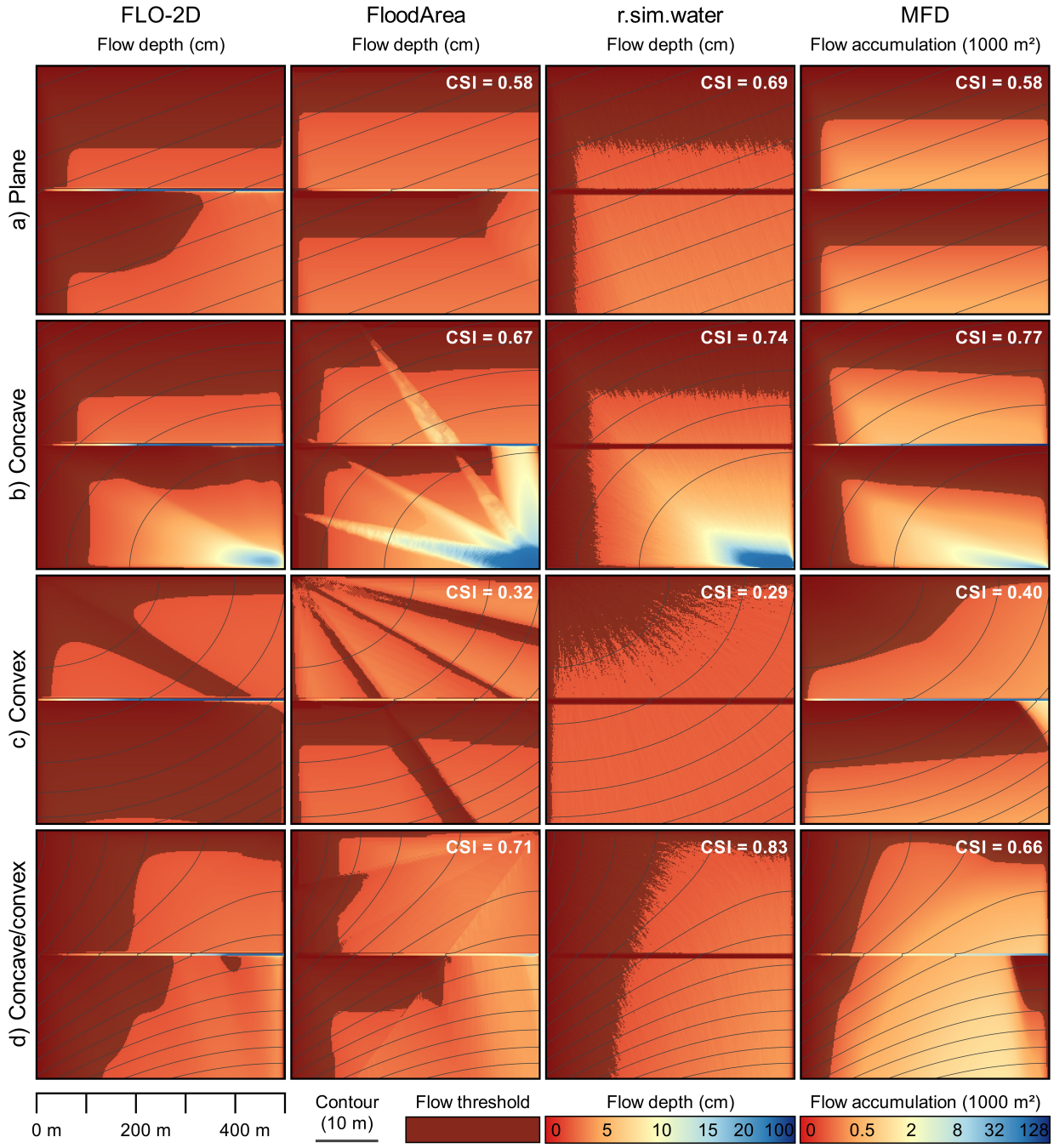


Figure 5: Simulation results of the four different models on a plane (a), concave (b), convex (c) and a combination of a convex and concave (d) artificial surface. The flow threshold for the hydrodynamic models (FLO-2D, FloodArea, r.sim.water) is a flow depth of 0.02 m, whereas the flow threshold for the flow accumulation algorithm (MFD) is an accumulation area of 250 m^2. Cells with values below the respective flow threshold are considered to be dry (dark red cells), while all other cells are considered to be wet. The indicated critical success index (CSI, cf. Eq. 10) was obtained by comparing the models' predicted wet and dry cells to the binary pattern produced by FLO-2D, which is used as a reference.

different from the other ones, which is reflected by the particularly low CSI values indicated in Fig. 5. Also the flow patterns in the northern half of the combined concave/convex surface seem to deviate slightly more among the models than the produced patterns on the plane and the concave surface, respectively. This could be explained by the fact that the northern half of the concave/convex surface is characterized by convex forms that produce particularly different results among the models. Lastly, the flow pattern produced by FloodArea on the concave surface is characterized by striking flow paths. FloodArea produces also sharp-edged flow paths on the convex and the concave/convex surface, however not as pronounced as on the concave surface. These flow patterns stem from the limitation of flow directions to 16 fixed angles by FloodArea's flow routing scheme, which is described in e.g. Tyrna et al. (2017).

3.2. Real-world case studies

The performance of the models applied to each case study is depicted in Fig. 6. The obtained CSI (Eq. 10) values are rather low and indicate that, overall, all models have a low performance for all case studies. The respective maximal CSI of each model lies between 0.318 and 0.344, which stem from the case study E2. For the same case study, the models produce the highest hit rates (Eq. 11), ranging between 0.566 and 0.788. Save a few exceptions, the hit rates are well below a value of 0.5 in all other case studies.

The bias is the fraction of simulated number of wet cells compared to the observed number of wet cells (Eq. 9). Thus, a bias greater than one indicates an overestimation of the wet cells by the model, whereas a bias below one shows the opposite. As presented in Fig. 6, all models overestimate the number of reconstructed wet cells for some case studies, but heavily underestimate them for others. As depicted in Fig. 6, the bias is correlated with the false alarm rates (Eq. 12). For each case study, models with a lower bias are also associated with a lower false alarm rate, and vice versa. The lowest absolute values are produced for the simulations with strong underestimations (bias $\ll 1.0$). This can be expected, since a particularly low bias value means that the number of modeled wet cells is much smaller than the observed number of wet cells. In this case, even if all modeled wet cells were misses, the false alarm rate would still be small, since the number of correct negatives is constantly high for all models. Hence, following Eq. 12, a low false alarm rate results.

Overall, we have identified three main issues limiting the models' performances, i.e., observational data of differing quality, insufficiently represented topographical structures and biased predictions of effective rainfall. In the following, we illustrate each of these issues with examples from the corresponding case studies.

The particularly low performance of all models applied to case study E1 can mainly be attributed to poor observational data. Namely, the derivation of observed wet cells are based on an external map (Jordi + Kolb AG, 2008). Therein, areas with ponded water as well as water on the streets are mapped, while overland flow paths in the agricultural fields are not indicated (Table 5). Thus, the observations only capture

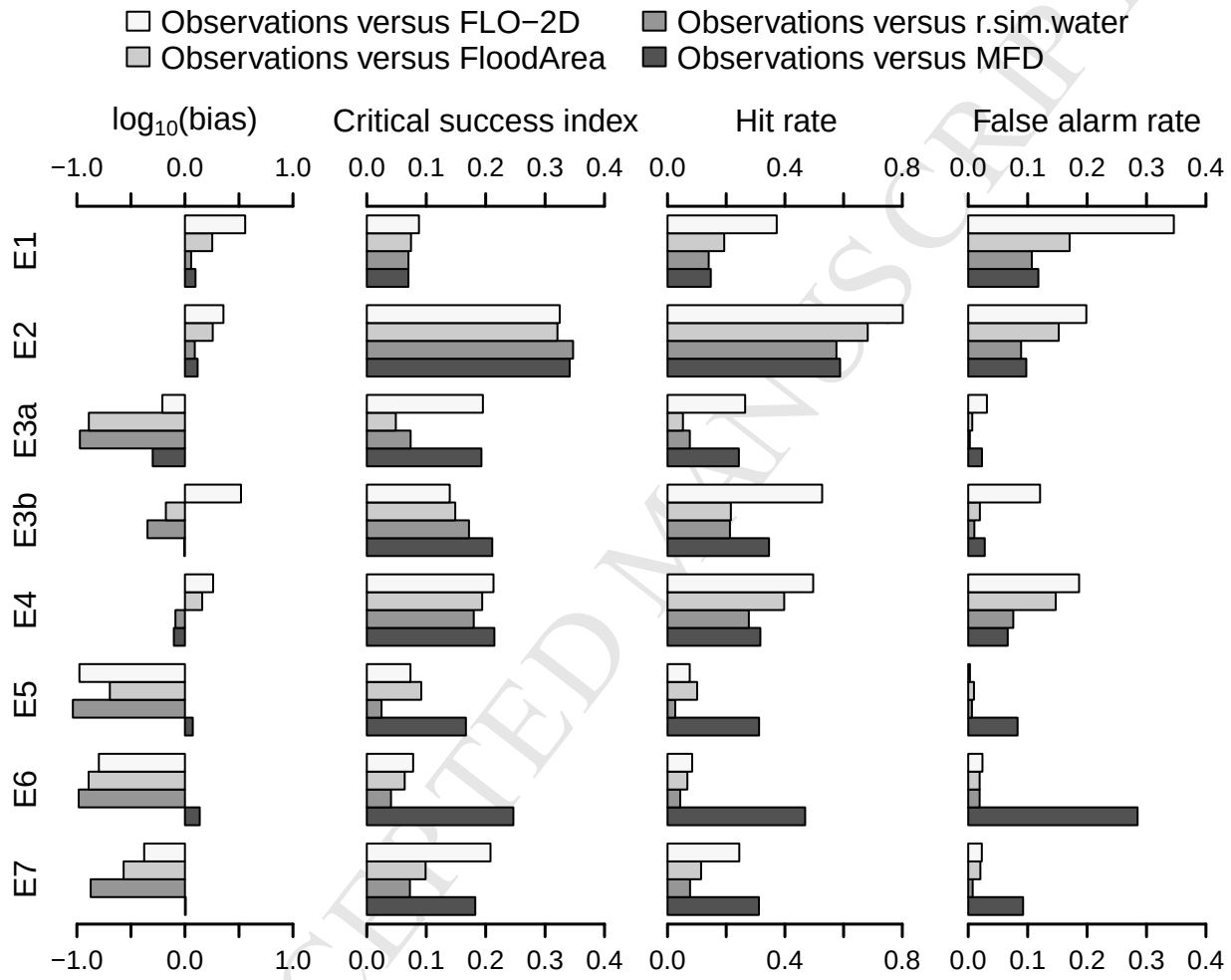


Figure 6: Model performance of each model in comparison to the observed values evaluated within the observation domain (D_{obs}, Sect. 2.3.1). The binary pattern performance measures are defined in Eq. 9-12. Note that not the bias itself, but the common logarithm of the bias is displayed. The IDs (E1-7) indicate the corresponding case study, as summarized in Table 2 and 3.

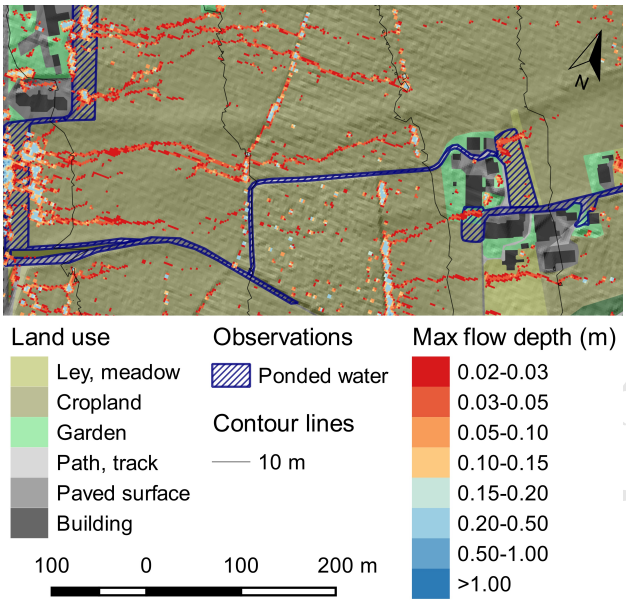


Figure 7: Excerpt of the simulation results produced by r.sim.water applied to the case study E1. The observations inferred from an external map only capture areas where water was ponding or where the street's drainage system was overwhelmed, indicated by the hatched blue areas. Flow paths in the agricultural areas were not mapped. The accumulating water along what looks like trenches in the central part of the figure, are in fact caused by artifacts of the DEM, as visualized by the transparent land use on top of the DEM's hillshade image. Note that the north direction is slightly tilted, as indicated.

areas that are small compared to the whole area that must have been inundated, as depicted by Fig. 7. As a consequence, the wet cells are overestimated, the false alarm rates are high and the CSI values are low for all models. Moreover, the map by Jordi + Kolb AG (2008) does not provide any ancillary information such as the applied mapping methods. Therefore, the map turns out to be an unsuitable source of information for the purpose of validating the models.

Applying the models to artificial surfaces has highlighted that topographical structures such as streets can have major effects on the produced flow paths (Sect. 3.1). How the models are predicting flow on streets in real-world case studies and how this influences the prediction of subsequent flow paths, can best be shown with results from the case study E2. All models perform best in this case study, as indicated by Fig. 6. The CSI values are similarly high for all models, whereas the other scores vary slightly more. For instance, FLO-2D produces the highest hit rate, however, at the expense of the highest false alarm rate and the highest overestimation. In contrast, r.sim.water exhibits the smallest bias and false alarm rate, however, at the expense of a smaller hit rate. Depending on the situation, one or the other configuration might be more desirable. The visual comparison confirms that all four models produce plausible results. As an example for the simulation results, the maximal flow depth predicted by FLO-2D are depicted in Fig. 8.

Based on the similar performance of all four models, the case study E2 is best suited for comparing the

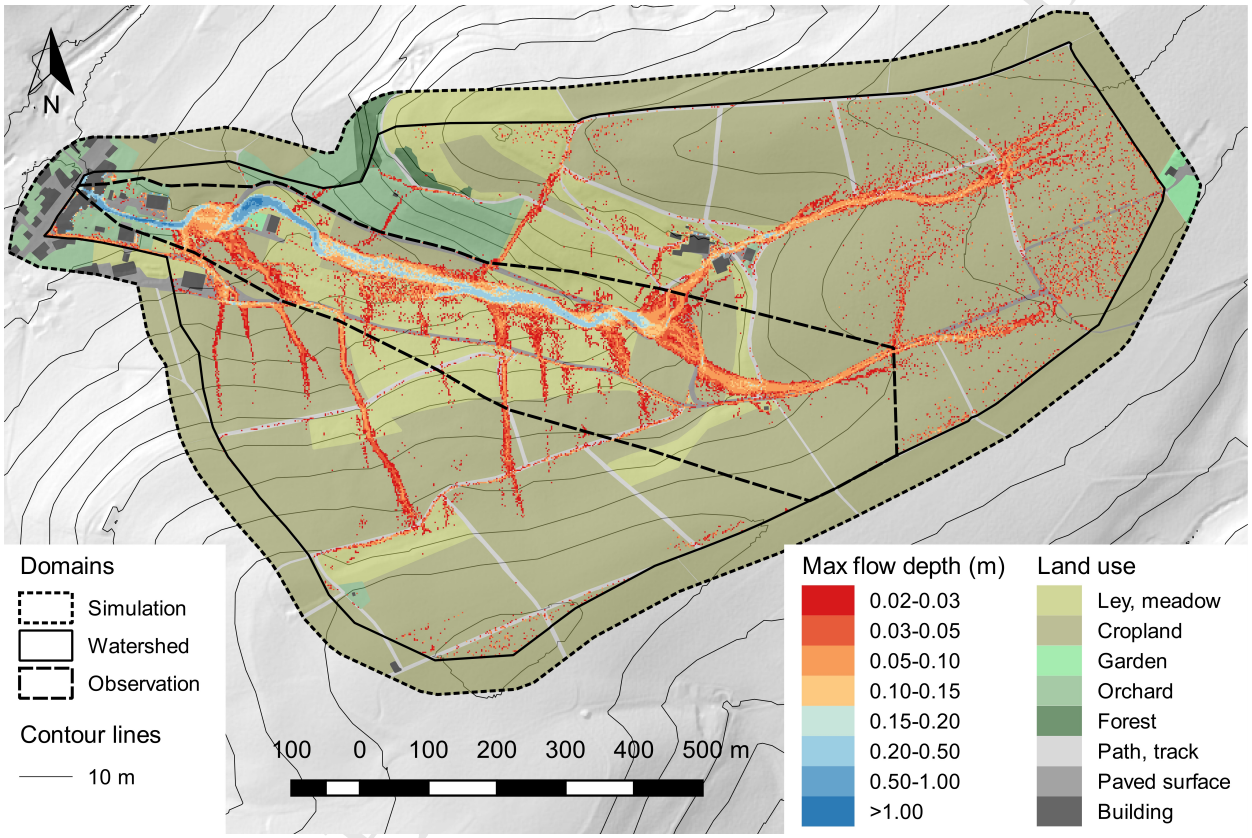


Figure 8: Simulation result of FLO-2D applied to the case study E2. The maximal flow depths are categorized into discrete classes, as indicated in the legend. Considering the chosen water depth threshold, all cells that display a maximal water depth of $d \geq 0.02$ m are simulated as wet, whereas all other cells are predicted to remain dry.

simulated flow paths in more detail. Namely, in most of the other case studies the models are associated with a greater range of bias values, i.e., the number of wet cells varies more among the models, which impairs the attribution of model differences. Fig. 9 illustrates the model comparison of the observed and simulated wet cells, as categorized according to the contingency table (Table 6).

According to all models, water mainly accumulates in the thalweg, i.e., the path of lowest elevations along the hillslope (cf. Fig. 8 and 9). Thereby, the observed wet cells are captured well by all models, except MFD, which is not able the predict the ponding water towards the outlet of the observation domain. The main differences between the other models are that FLO-2D is overestimating the wet cells along the thalweg more than FloodArea, which in turn overestimates the wet cells to a larger degree than r.sim.water. This is reflected by the respective bias values indicated in Fig. 9.

Overall, the models have difficulties predicting the water flowing on the streets. Foremost, the streets in the upper part of the domain were inundated, but were not simulated as such, which is reflected by the numerous misses in this area (red cells, Fig. 9). Compared to the other models, FLO-2D predicts the observed wet street cells better. This behavior could be expected based on the results from the models applied to artificial surfaces, since FLO-2D predicted consistently larger flow depths on the street than the other models (Sect. 3.1). Along the same lines, r.sim.water predicts the lowest number of wet street cells, which is also supported by the findings of the artificial modeling exercise. Interestingly, all models predict roughly the same places where water overtops the street's confinement and joins the main flow path in the thalweg. Only one of these paths in the central part of the domain is predicted by FloodArea and FLO-2D, while the path is not indicated by r.sim.water and MFD. East thereof, a flow path could be observed that is not simulated by any model. Overall, this exemplifies that although the behavior may differ slightly between the models on a cell-by-cell basis, they all produce quite similar flow paths on a broader scale or, similarly fail to identify them.

As outlined introductorily, accurate predictions of effective rainfall are crucial for increased model performances, in addition to high-quality observational data and well-represented topographical structures. If a model predicts too little runoff, it usually leads to an underestimation of wet cells and, consequently, to a rather low performance. This issue is nicely exemplified by the case studies E3a and E3b observed at the same study site (cf. Table 1). As shown in Fig. 6, the number of wet cells are underestimated by all models in the first event (E3a). In particular, FloodArea and r.sim.water predict a much lower number of wet cells than the number of wet cells inferred from the observations. Consequently, the performance of these two models is particularly low for this case study. The performances are more balanced for the second observed event (E3b). However, similarly to the case study E2, FLO-2D produces the highest hit rate, but also the highest false alarm rate, owed to the overestimated number of wet cells. Although, both case studies were triggered by thunderstorms, the rainfall intensities of E3a are moderate and the event spans 13 hours, while E3b is associated with short and intense rainfall that is typical for thunderstorms (cf. Table 2).

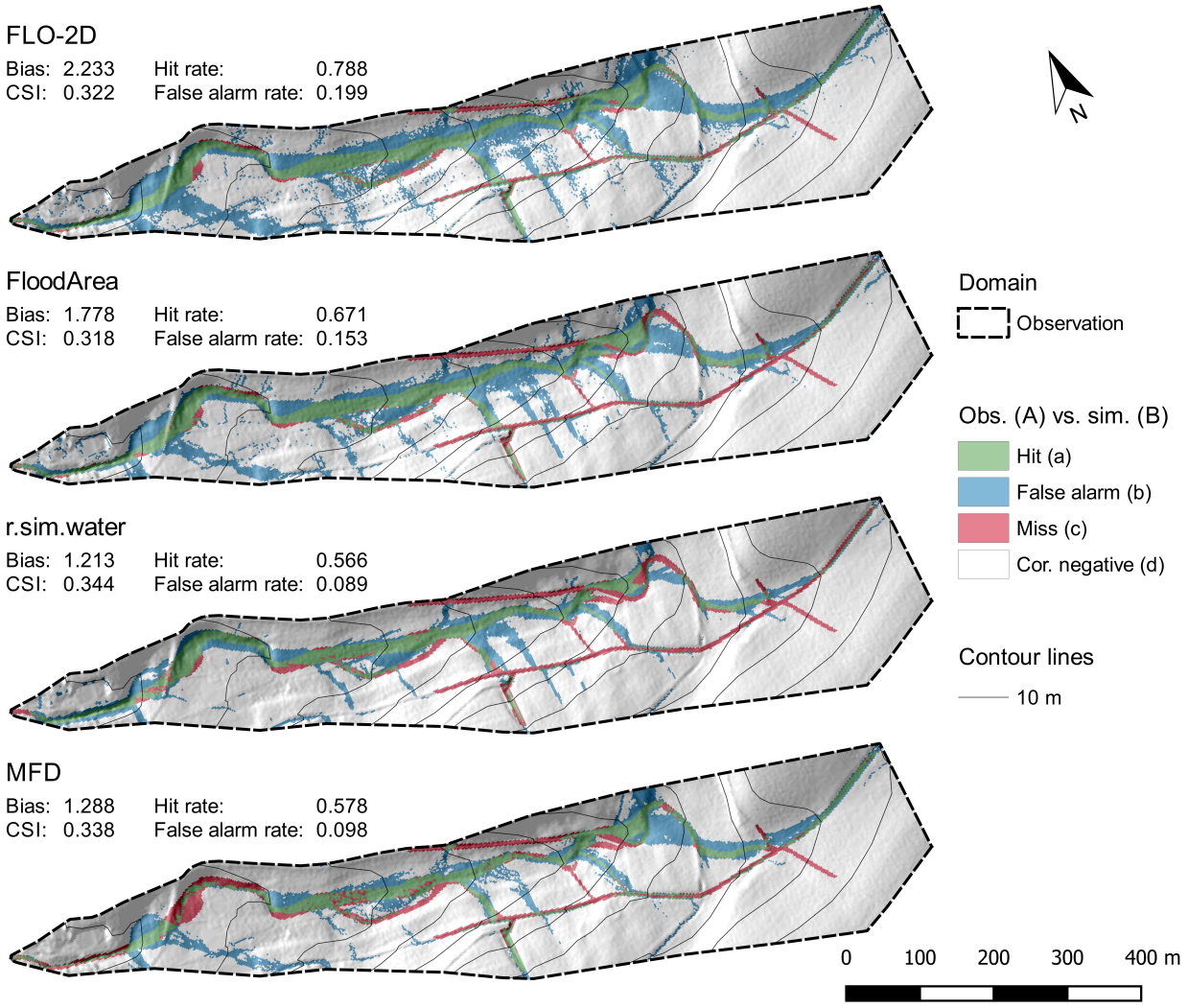


Figure 9: Comparison of observed (obs.) and simulated (sim.) wet cells as categorized by the contingency table (Table 6) for each model applied to the case study E2. The definitions of hits, false alarms, misses and correct negatives can be found in Table 6. The whole study area (cf. Fig. 8) is clipped to the observation domain, as observations are unavailable outside of this domain. Note that the north direction is slightly tilted.

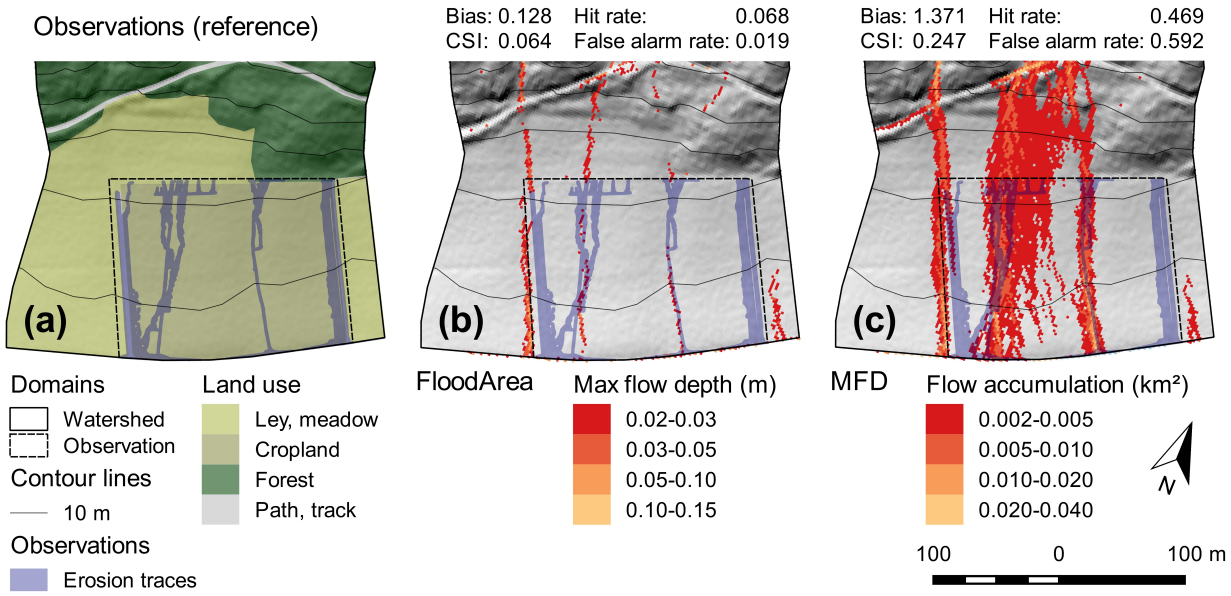


Figure 10: Exemplary comparison of inundated areas inferred from documented traces of with flow depths predicted by FloodArea and flow accumulation calculated with MFD for the case study E6. (a) Flow paths inferred by means of a UAV in a patato field (cf. Fig. 4a). The land use is displayed in the background. (b) Maximal flow depth simulated by FloodArea. (c) Flow accumulation as predicted by MFD. Note that the north direction is slightly tilted for all sub-figures.

Thus, we can observe that the hydrodynamic models, i.e., all except MFD, generally underestimate the number of wet cells for the case studies with low rainfall intensities. Namely, the said models exhibit an underestimation of the observed wet cells for the case studies E3a, E5, E6 and E7, as depicted in Fig. 6. This hints at the fact that the simulation of wet cells is less sensitive for case studies driven by intensive rainfall. In contrast, the mechanisms that lead to overland flow during the case studies with low rainfall intensities, are much more complex and badly captured by the chosen modeling approach of this study.

The model MFD is inherently different from the other three hydrodynamic models. It is not an event-based model, but assesses a static property of a catchment based solely on the DEM, i.e., the relative catchment area (Sect. 2.1.4). Applied to the considered case studies, MFD performs similarly or even better than the other models. This is most pronounced in the case study E6, further illustrated in Fig. 10.

As highlighted by Fig. 10, the number of wet cells predicted by FloodArea is particularly low for the case study E6 within the observation domain, which is reflected by the performance measures' low values (Fig. 10b). Although the flow paths in the middle of the observation domain are vaguely indicated, it is apparent that too little effective rainfall is predicted, which leads to the exhibited underestimation of wet cells. In contrast, the flow paths predicted by MFD are a function of the respective catchment area of each cell, irrespective of the rainfall. Thereby, the two flow paths in the middle of the observation domain are covered. Notably, the flow path at the western border of the observation domain is shifted slightly westwards

in comparison to the observations. This behavior can be attributed to the specific land management of the corresponding patato field, i.e., furrows parallel to the slope, which promotes flow at western border of this field. The observed flow path at the eastern border of the observation domain is not captured by any model. Moreover, it should be noted that the flow over the bare potato field led to erosion, as depicted in Fig. 4a, which in turn may have a significant influence on the flow patterns. However, such effects cannot be captured with this study's modeling approach.

Since in most case studies the observations only cover a rather small part of the whole simulated catchment (cf. D_{obs} and D_{wsd} in Table 2-3) and the observations are associated with a varying degree of confidence (cf. Table 5), we additionally compare the model results within the whole simulation domain with each other independent of the observation data. By using the more sophisticated models as the reference, we can assess the capability of the simpler models to reproduce results of the more complex models. As Fig. 11 indicates, FloodArea as well as MFD reproduce the results stemming from FLO-2D rather well. At the same time, the false alarm rates are particularly low. However, we also recognize that FloodArea generally underestimates the wet cells in comparison to FLO-2D. r.sim.water slightly underestimates the wet cells in comparison to FloodArea. However, the underestimation is limited to a small range indicating that the underestimation is similar in all case studies.

In addition to the comparison of the models among themselves, the first row and column of Fig. 11 also depicts the model performance in relation to the observations. Namely, it also displays the results shown in Fig. 6 in a different way, whereby the overall performance is better visualized. Thus, it depicts that the CSI of every model is rather low, as discussed before. Moreover, it visualizes that the CSI of the model MFD and FLO-2D is very similar, as well as the one of FloodArea and r.sim.water, but at a lower level. Moreover, it visualizes the stronger tendency of FloodArea and r.sim.water to underestimate the wet cells, compared to FLO-2D. In comparison, MFD is by far the least biased of all the models.

4. Discussions

In this study, we have followed the procedure employed in practice by current hazard assessments to produce SWF hazard maps, which are based on uncalibrated, single deterministic simulations (cf. Meon et al., 2009; Tyrna & Hochschild, 2010; Kipfer et al., 2015; Tyrna et al., 2017). The results from the models applied to artificial surfaces and eight real-world case studies suggest that the models' performance might be increased if the model were properly calibrated. For instance, the model exercise on artificial surfaces (cf. Sect. 3.1) exemplified the need to calibrate the surface roughness. Namely, FLO-2D predicts rather high flow depths on the incised street, while r.sim.water predicts flow depth that are even below the chosen flow threshold of 0.02 m. Simulations with altered roughness values indicated that r.sim.water is rather sensitive to the street's chosen roughness value. Similarly, the flow depths on the street predicted by FloodArea are

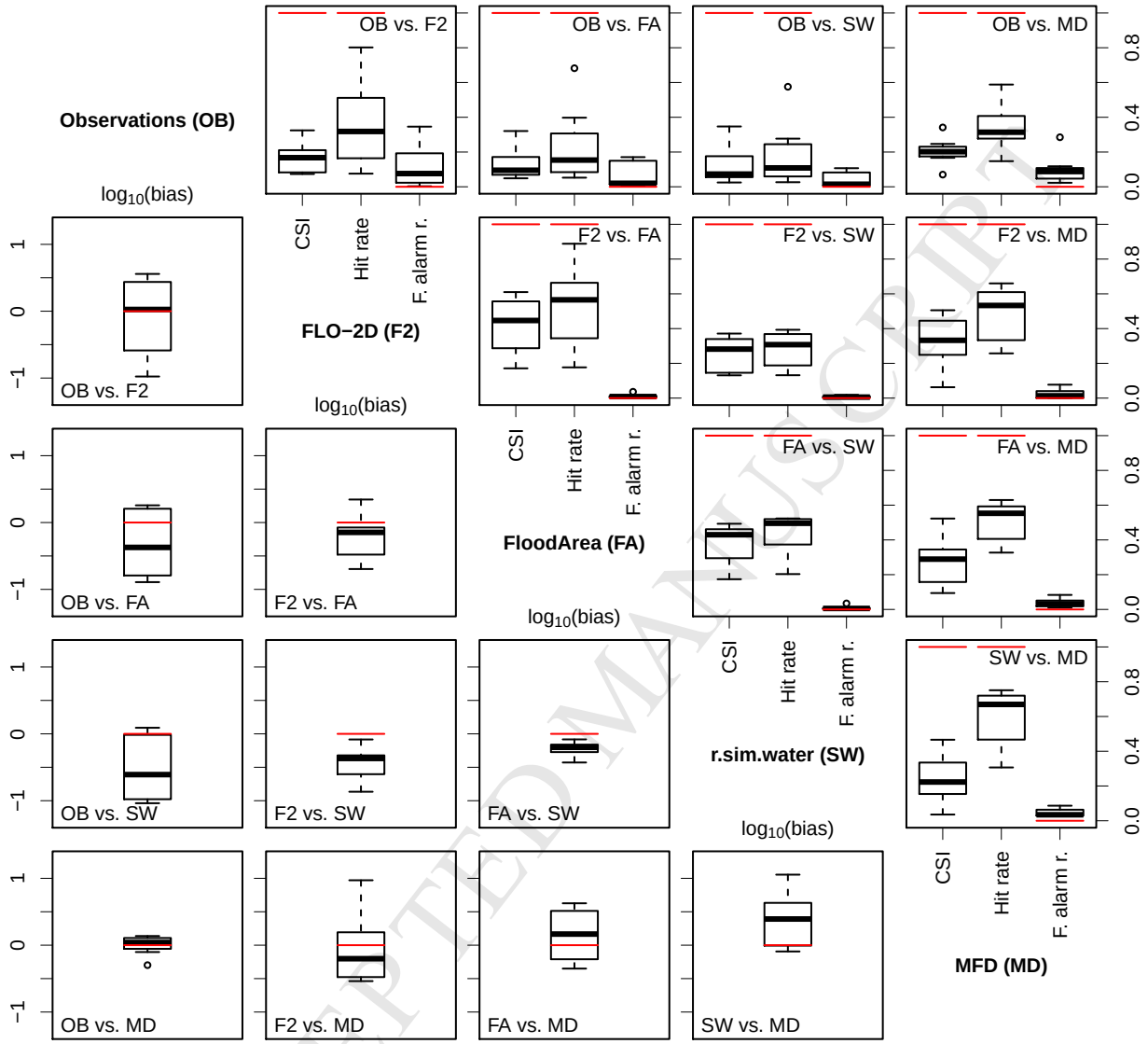


Figure 11: Binary pattern performance measures of different pair-wise model and/or observation comparisons. Each box plot is built by eight values, i.e., one value for each case study. In the first row and the first column, the observations are compared with the model results within the observation domain (D_{obs}, Sect. 2.3.1). All other sub-figures show the comparison of different models within the watershed domain (D_{wsd}, Sect. 2.3.1). Each sub-figure is labeled with the abbreviated pairing, whereas the former label indicates the reference to which the latter is compared. As an example, the label "OB vs. SW" indicates the sub-figure, in which the observations are compared to the model results of r.sim.water. In the sub-figures above the diagonal, the performance measures are plotted, i.e., the critical success index (CSI, cf. Eq. 10), the hit rate (cf. Eq. 11) and the false alarm rate (f. alarm r., cf. Eq. 12). The bias (cf. Eq. 9) is plotted in the sub-figures below the diagonal. Note that not the bias itself, but the common logarithm of the bias is displayed. The ideal value of each performance measure (cf. Eq. 9-12) is indicated by the thin red line. The closer the box plots are to this red line, the more similar are the performance measures of the corresponding pairing.

generally below those predicted by FLO-2D. A calibration of the roughness value could also improve the match between FLO-2D and FloodArea. This circumstance is also exhibited by applying the models in real-world case studies, whereby FloodArea and r.sim.water predict lower flow depths on streets (cf. Fig. 9 and Sect. 3.2).

The results from applying the models to a broad range of different settings indicate that the models' performance would still vary significantly, even if the models were calibrated. Namely, all models perform similarly well in the case study E2 (cf. Fig. 6), whereas properly calibrated models might perform even better. Yet, it is clear that a calibration could not bring the models' performances to a similar level in all case studies. On the one hand, this indicates that calibration and/or validation based on one single case study might be misleading. Thus, using various case studies covering a wide range of settings provides a more holistic picture of the models' performance. On the other hand, it also indicates that the models are not capable of capturing all relevant processes under diverse circumstances.

In fact, the results show that the hydrodynamic models tend to significantly underestimate the number of wet cells for the case studies associated with weak rainfall (cf. Fig. 6 and 10). Thus, the models do not predict sufficient runoff as compared to the observations, driven by an underestimation of the effective rainfall. More specifically, the results indicate that the considered infiltration assessment methods (cf. Sect. 2.3.5) are not not capturing the governing processes well. Namely, saturation excess overland flow cannot be modeled by the applied methods, although this runoff generation mechanism is likely crucial for SWFs triggered by weak rainfall. Although SWFs are usually associated by heavy rainfall as mentioned before, results from Bernet et al. (2017) indicate that long lasting events with weak rainfall cause similar damage to buildings as short events with heavy rainfall. Thus, a model should be able to capture events characterized by heavy as well as weak rainfall to be suitable to reliably simulate SWFs in rural areas.

Along these lines, the two events observed at the same study site, i.e., case study E3a and E3b, exemplify that SWFs can be triggered by heavier and weaker rainfall at the same location (cf. Table 2). Moreover, the case studies exemplify that the flow paths are not a static function of the topography, but are dependent on soil characteristics, land use, land management in addition to the rainfall input, of course. Along these lines, Ferreira et al. (2015) highlighted for instance that runoff generation mechanism are spatially and temporally highly variable. Certainly, there are established and emerging methods that could represent the runoff generation processes better (e.g. Schmocker-Fackel et al., 2007; Antonetti et al., 2016; Steinbrich et al., 2016). However the consideration of such spatially highly variable processes are often impaired by the lack of appropriate data. Thus, for a better representation of the runoff generation processes, corresponding data are required. For the presented case studies such data were unavailable, as well as time-consuming and costly to collect.

The representation of topographical structures by the DEM is another aspect, which significantly influences the models' predictions (e.g. Sampson et al., 2012; de Almeida et al., 2016). As indicated by the

model exercise on artificial surfaces, the models react sensitively on structures such as streets (Sect. 3.1). Moreover, applying the models to real-world case studies have pinpointed that the influence of such structures on the simulation results are governed more by the representation of such structures by the respective DEM than the choice of the model by the user. This is in line with findings stemming from more formal model comparisons, for instance from the benchmark study of urban flood models by Fewtrell et al. (2011). This issue is illustrated by Fig. 9, which indicates that the models predict the streets' overtopping at the same locations, while numerous of these flow paths could not be observed in reality. Thus, this behavior suggests that the streets confinements are not represented accurately enough by the DEM, supported by the fact that the rural environment of the case study E2 is characterized by single-lane streets with width in the same order as the DEM's resolution. In consequence, the channelizing effect of overland flow on streets is rather poorly captured by the models. Confronted with the same issue, Kipfer et al. (2012) proposed to incise all streets by a fixed depth. However, this measure most likely incapacitate the model to correctly reproduce the street's overtopping. Thus, a more common solution is to use a DEM with a finer resolution, if available (Wechsler, 2007; Dottori et al., 2013). Generally, small-scale structures such as narrow streets are certainly better represented by a DEM with finer resolutions (Wechsler, 2007; Fewtrell et al., 2011; de Almeida et al., 2016). Nevertheless, as has been pointed out before, finer resolutions might also lead to inadequate confidence in the extremely precise model outputs (Dottori et al., 2013). Along these lines, it is crucial to note that the DEM itself is an imperfect representation of the reality, irrespective of its resolution (Wechsler, 2007; Abily et al., 2016). Just as DEMs with coarser resolutions, topographical models with finer resolutions are not flawless either and contain artifacts, which may cause false results, as illustrated in Fig. 7. Therefore, the DEM needs to receive particular attention, i.e., it needs to be carefully pre- and post-processed in order to represent realistic flow paths, as other studies have highlighted as well (Hankin et al., 2008; Hunter et al., 2008; Tyrna et al., 2017).

The models may produce distinctly different results under certain circumstances, as highlighted by the models applied to the convex artificial surface characterized by diverging flow patterns (Fig. 5). However, in real-world applications such forms are likely less important in comparison to plane and concave slopes, as exemplified by the case study E2 characterized by concave topography (cf. Fig 8). On such slopes, the models exhibit better model agreement (Fig. 5). Thus, in real-world applications, the model choice seems to play an inferior role compared to the previously discussed issues including the appropriate representation of effective rainfall and topographical structures. In other words, the model choice is generally not the most important factor determining whether the observed inundation area can be predicted well by the corresponding model, at least for events associated with heavy rainfall. However, it should be noted that this statement might be different for the prediction of flow depths and/or flow velocities. As mentioned before, the hydrodynamic models, i.e., FLO-2D, FloodArea and r.sim.water, generally predict the number of wet cells less reliably for case studies associated with weak rainfall. An exception is the flow algorithm MFD, which produces the

least biased results for all case studies (Fig. 11). At the downside, the algorithm cannot consider ponding or backwater, as exemplified by the model's distinct underestimation of the inundated area towards the outlet of the study site E2 (Fig. 9, MFD). Thus, MFD can be used for approximating the extent of inundated areas, but not for predicting flow depths, flow velocities, and flow dynamics in general.

Lastly, the model performance is also highly dependent on the used data. Therefore, it is crucial to account for the uncertainties introduced by the input data, for instance by carrying out a sensitivity analysis (e.g. Pianosi et al., 2016). At the same time, the uncertainties need to be considered, which stem from the observational data that are used to condition and/or evaluate the models. For instance, if the observational data are a bad representation of the models' simulated quantity, the performance of the models are inevitably low, as exemplified by the case study E1 (Fig. 7 and Sect. 3.2). Yet, as mentioned before, there is little high-quality observational data available, which exhibit appropriate spatial and temporal resolutions suitable for model calibration and/or validation (e.g. Hunter et al., 2008; Blanc et al., 2012; Neal et al., 2012; Yu & Coulthard, 2015). Consequently, in this study, it was necessary to exploit different data sources, including external maps, eye witnesses' photographs and videos, mapped flood traces based on field visits partly supported by aerial photographs (cf. Table 5 and Fig. 4). Yet, the mapped quantity is not the same for each source. While the exploited photographs and videos represent a snapshot of the flow pattern at a certain instant during the respective SWF, reconstructions based on flood marks are constrained to areas where the flood has left discernible traces. For instance, overland flow with few suspended particles might not leave identifiable traces. In consequence, this likely leads to an underestimation of the actual inundated area. Accordingly, we assigned this data source with a lower (medium) confidence level, as indicated in Table 5. Despite the increased confidence level for overland flow reconstructed from photographs and videos, a similar bias might apply to this data source, as well. Namely, a bias is introduced if the picture is not taken at the instant of the maximal flood extent.

Therefore, just as the simulation outputs, the observational data should be regarded as uncertain (Bennett et al., 2013; Stephens et al., 2014; Savage et al., 2016a). Thus, the representation of the observations and simulations as a crisp representation of the reality might be inappropriate. To address this issue, Pappenberger et al. (2007) applied a fuzzy set approach to measure the performance based on uncertain observational data. Thereby, slight shifts between observed and simulated wet cells could be accounted for. For simulated wet cells, it is straightforward to obtain a confidence level that a particular cell is wet by considering the simulated flow depths (Pappenberger et al., 2007). In contrast, this is not trivial for the observational data used in this study. Namely, ancillary data would be necessary. For instance, flood traces mapped in the field could be categorized according to the respective confidence that the corresponding area was in fact inundated.

In case only the flood extent is of interest, for instance when identifying potentially flooded assets, choosing a simple over a hydrodynamic model might be advantageous: as exemplified by the real-world case

studies, the extent of SWFs can be predicted similarly well with MFD as with the considered hydrodynamic models, while the associated computational demand is much smaller. Consequently, MFD could be applied to larger areas while exploiting the increasing availability of high-resolution DEMs. Moreover, there are other simple conceptual models, as termed by Teng et al. (2017), which may overcome some of the limitations of MFD, such as the incapability to simulate ponding water. Potential candidates include for instance the Rapid Flood Spreading Method (RFSM), as described by L'homme et al. (2008), or the model called HAND (height above the nearest drainage), as introduced by Nobre et al. (2011). Such approaches could be applied to (almost) any scale and area (Teng et al., 2017), which could make them interesting candidates for regional or even continental hazard assessments regarding SWFs. Moreover, such models are predestined to be used in probabilistic modeling approaches (e.g. Merwade et al., 2008; Aronica et al., 2012; Savage et al., 2016a). Thus, the applicability of a probabilistic modeling approach in relation to SWFs in rural areas should be investigated in the future, as well.

5. Conclusions and outlook

The main aim of this study was to test a SWF hazard assessment approach that is currently employed in practice and is based on single simulations with uncalibrated and/or unvalidated flood inundation models. For that matter, we applied four uncalibrated raster-based models to four characteristic artificial surfaces and eight real-world case studies. The models' application to the artificial surfaces exemplified that the flow patterns are heavily disturbed by streets, insofar as the prediction of inundated areas downslope of such structures are significantly influenced. Thus, there are large differences of how each model predicts these flow disturbances. Moreover, the modeling exercise has indicated that the models disagree most about the prediction of flow on the convex surface. The performance of the models applied to real-world case studies was assessed qualitatively as well as quantitatively in relation to inundated areas inferred from different sources. In summary, the performance of the selected grid-based models indicates that they are not (yet) suited to be employed in an uncalibrated mode to reliably and deterministically predict inundated areas caused by SWFs in rural areas. Mainly, the models' performances are impaired by biased predictions of effective rainfall and insufficient representation of topographical structures.

To improve the prediction of SWF hazards, various approaches seem prospective. First of all, the deterministic modeling approach could be improved by incorporating a better prediction of the complex runoff generation mechanisms under various conditions. Moreover, the representation of topographical structures could be improved by considering DEMs with finer resolutions. Alternatively, irregular meshes and corresponding models could be used for a better representation of structures such as streets. At the same time, this study indicates that the models' calibration and/or their results' validation is imperative. For this task, the uncertainties of the observations should be considered, which may vary significantly depending on

the source and quality of the observations. In general, the quantification and communication of the models' associated uncertainties are crucial, as the models' extremely precise outputs have indeed the potential to provoke overconfidence in their results, which may lead to inappropriate decisions in flood risk management (Dottori et al., 2013).

A different way forward would be to exploit simple conceptual models such as MFD. Within the context of this study, MFD performed similarly well than the hydrodynamic models. Thus, similar conceptual models could be tested, which overcome some of the limitations of MFD, while providing similar results. The computational effort of such simple models is by far the least. Such approaches are therefore also interesting for the application to large areas, for instance in the context of regional, national or even continental SWF hazard assessments. Yet, due to lower computational constraints, even the topographical data with the finest available resolutions might be exploited. Moreover, such models could be applied in a probabilistic simulation framework, which could potentially better handle the lack of observational data in comparison to the current deterministic approaches.

Finally, this study highlighted once more that observational data are crucial irrespective of the chosen way forward. Thus, a standardized method to document and report SWFs in rural and urban areas is required and should be developed. At the same time, systematic observations should be put in place to lie the ground for future research, which is certainly necessary.

Acknowledgments

Funding from the Swiss Mobiliar supported the completion of this research. We thank the Federal Office of Topography for providing the corresponding spatial data, as well as the Federal Office of Meteorology and Climatology for providing the rainfall data. We would also like to thank geomer GmbH for providing a FloodArea license under favorable terms, and in particular Andr Assmann for his assistance. Moreover, we thank Mirjam Stawicki for supporting the field work, David Thöni for his modeling efforts during the initial phase of the study, Simona Trefalt for proofreading, and Guido Felder for revising the manuscript.

References

Abily, M., Bertrand, N., Delestre, O., Gourbesville, P., & Duluc, C.-M. (2016). Spatial global sensitivity analysis of high resolution classified topographic data use in 2D urban flood modelling. *Environ. Modell. Softw.*, *77*, 183–195. doi:10.1016/j.envsoft.2015.12.002.

Alder, S., Prasuhn, V., Liniger, H., Herweg, K., Hurni, H., Candinas, A., & Gujer, H. U. (2015). A high-resolution map of direct and indirect connectivity of erosion risk areas to surface waters in Switzerland — A risk assessment tool for planning and policy-making. *Land Use Policy*, *48*, 236–249. doi:10.1016/j.landusepol.2015.06.001.

de Almeida, G. A. M., Bates, P., & Ozdemir, H. (2016). Modelling urban floods at sub-metre resolution: challenges or opportunities for flood risk management? *J. Flood Risk Manage.*, (pp. 1–11). doi:10.1111/jfr3.12276.

Andrieu, H., Browne, O., & Laplace, D. (2004). Les crues en zone urbaine: des crues éclairs? *La Houille Blanche*, (pp. 89–95). doi:10.1051/lhb:200402010.

Antonetti, M., Buss, R., Scherrer, S., Margreth, M., & Zappa, M. (2016). Mapping dominant runoff processes: An evaluation of different approaches using similarity measures and synthetic runoff simulations. *Hydrol. Earth Syst. Sc.*, *20*, 2929–2945. doi:10.5194/hess-20-2929-2016.

Aronica, G., Bates, P. D., & Horritt, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrol. Process.*, *16*, 2001–2016. doi:10.1002/hyp.398.

Aronica, G. T., Franza, F., Bates, P. D., & Neal, J. C. (2012). Probabilistic evaluation of flood hazard in urban areas using Monte Carlo simulation. *Hydrol. Process.*, *26*, 3962–3972. doi:10.1002/hyp.8370.

Barredo, J. I. (2009). Normalised flood losses in Europe: 1970–2006. *Nat. Hazards Earth Syst. Sci.*, *9*, 97–104. doi:10.5194/nhess-9-97-2009.

Bennett, N. D., Croke, B. F., Guariso, G., Guillaume, J. H., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. *Environ. Modell. Softw.*, *40*, 1–20. doi:10.1016/j.envsoft.2012.09.011.

Bernet, D. B., Prasuhn, V., & Weingartner, R. (2017). Surface water floods in Switzerland: what insurance claim records tell us about the damage in space and time. *Nat. Hazards Earth Syst. Sci.*, *17*, 1659–1682. doi:10.5194/nhess-17-1659-2017.

Blanc, J., Hall, J. W., Roche, N., Dawson, R. J., Cesses, Y., Burton, A., & Kilsby, C. G. (2012). Enhanced efficiency of pluvial flood risk estimation in urban areas using spatial-temporal rainfall simulations. *J. Flood Risk Manage.*, *5*, 143–152. doi:10.1111/j.1753-318X.2012.01135.x.

Castro, D., Einfalt, T., Frerichs, S., Friedeheim, K., Hatzfeld, F., Kubik, A., Mittelstädt, R., Müller, M., Seltmann, J., & Wagner, A. (2008). *Vorhersage und Management von Sturzfluten in urbanen Gebieten (URBAS): Schlussbericht des vom Bundesministerium für Bildung und Forschung geförderten Vorhabens*. Aachen, Deutschland: Hydrotec GmbH and Fachhochschule Aachen and Deutscher Wetterdienst. URL: http://www.urbanesturzfluten.de.

CEPRI (2014). *Gérer les inondations par ruissellement pluvial — Guide de sensibilisation*. Orléans, France: Centre Européen de Prévention du Risque d'Inondation. URL: http://www.cepri.net/Ruissellement_pluvial.html.

Chan, F. K. S., Mitchell, G., Adekola, O., & McDonald, A. (2012). Flood risk in Asia's urban mega-deltas. *Environment and Urbanization ASIA*, *3*, 41–61. doi:10.1177/097542531200300103.

Chen, A. S., Evans, B., Djordjević, S., & Savić, D. A. (2012). A coarse-grid approach to representing building blockage effects in 2D urban flood modelling. *J. Hydrol.*, *426–427*, 1–16. doi:10.1016/j.jhydrol.2012.01.007.

Chu, S. T. (1978). Infiltration during an unsteady rain. *Water Resour. Res.*, *14*, 461–466. doi:10.1029/WR014i003p00461.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.*, *8*, 1991–2007. doi:10.5194/gmd-8-1991-2015.

Cook, A., & Merwade, V. (2009). Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. *J. Hydrol.*, *377*, 131–142. doi:10.1016/j.jhydrol.2009.08.015.

Coulthard, T. J., & Frostick, L. E. (2010). The Hull floods of 2007: implications for the governance and management of urban drainage systems. *J. Flood Risk Manage.*, *3*, 223–231. doi:10.1111/j.1753-318X.2010.01072.x.

Cutter, S. L., & Emrich, C. (2005). Are natural hazards and disaster losses in the U.S. increasing? *Eos Trans. AGU*, *86*, 381–389. doi:10.1029/2005EO410001.

Di Baldassarre, G., Montanari, A., Lins, H., Koutsoyiannis, D., Brandimarte, L., & Blöschl, G. (2010). Flood fatalities in Africa: From diagnosis to mitigation. *Geophys. Res. Lett.*, *37*, L22402. doi:10.1029/2010GL045467.

Dottori, F., Di Baldassarre, G., & Todini, E. (2013). Detailed data is welcome, but with a pinch of salt: Accuracy, precision,

and uncertainty in flood inundation modeling. *Water Resour. Res.*, *49*, 6079–6085. doi:10.1002/wrcr.20406.

Dunkerley, D. (2008). Identifying individual rain events from pluviograph records: A review with analysis of data from an Australian dryland site. *Hydrol. Process.*, *22*, 5024–5036. doi:10.1002/hyp.7122.

DWA (2013). *Starkregen und urbane Sturzfluten — Praxisleitfaden zur Überflutungsvorsorge* volume T1/2013 of *DWA-Themen*. Hennef, Deutschland: Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA).

Egli, T. (2007). *Wegleitung Objektschutz gegen meteorologische Naturgefahren*. Bern, Schweiz: Vereinigung Kantonaler Feuerversicherungen. URL: http://vkf.ch/VKF/Downloads.

Falconer, R. H., Cobby, D., Smyth, P., Astle, G., Dent, J., & Golding, B. (2009). Pluvial flooding: new approaches in flood warning, mapping and risk management. *J. Flood Risk Manage.*, *2*, 198–208. doi:10.1111/j.1753-318X.2009.01034.x.

Ferreira, C., Walsh, R., Steenhuis, T. S., Shakesby, R. A., Nunes, J., Coelho, C., & Ferreira, A. (2015). Spatiotemporal variability of hydrologic soil properties and the implications for overland flow and land management in a peri-urban Mediterranean catchment. *J. Hydrol.*, *525*, 249–263. doi:10.1016/j.jhydrol.2015.03.039.

Fewtrell, T. J., Bates, P. D., Horritt, M., & Hunter, N. M. (2008). Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrol. Process.*, *22*, 5107–5118. doi:10.1002/hyp.7148.

Fewtrell, T. J., Duncan, A., Sampson, C. C., Neal, J. C., & Bates, P. D. (2011). Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data. *Phys. Chem. Earth*, *36*, 281–291. doi:10.1016/j.pce.2010.12.011.

Foody, G. M. (2007). Map comparison in GIS. *Prog. Phys. Geog.*, *31*, 439–445. doi:10.1177/0309133307081294.

Freeman, T. G. (1991). Calculating catchment area with divergent flow based on a regular grid. *Comput. Geosci.*, *17*, 413–422. doi:10.1016/0098-3004(91)90048-I.

Fuchs, S., Keiler, M., & Zischg, A. (2015). A spatiotemporal multi-hazard exposure assessment based on property data. *Nat. Hazards Earth Syst. Sci.*, *15*, 2127–2142. doi:10.5194/nhess-15-2127-2015.

Fuchs, S., Röthlisberger, V., Thaler, T., Zischg, A., & Keiler, M. (2017). Natural hazard management from a coevolutionary perspective: Exposure and policy response in the European Alps. *Ann. Am. Assoc. Geogr.*, *107*, 382–392. doi:10.1080/24694452.2016.1235494.

Gaitan, S., van de Giesen, N. C., & ten Veldhuis, J. A. E. (2016). Can urban pluvial flooding be predicted by open spatial data and weather data? *Environ. Modell. Softw.*, *85*, 156–171. doi:10.1016/j.envsoft.2016.08.007.

geomer (2016). *FloodAreaHPC-Desktop: ArcGIS-Extension for calculating flooded areas — User manual: Version 10.3*. Heidelberg, Germany: geomer GmbH and Ruiz Rodriguez+Zeisler+Blank. URL: http://www.geomer.de/fileadmin/templates/main/res/downloads/UserManualFloodArea10.pdf.

Grahn, T., & Nyberg, L. (2017). Assessment of pluvial flood exposure and vulnerability of residential areas. *Int. J. Disaster Risk Reduct.*, *21*, 367–375. doi:10.1016/j.ijdrr.2017.01.016.

Green, W. H., & Ampt, G. A. (1911). Studies on soil phyics. *J. Agric. Sci.*, *4*, 1. doi:10.1017/S0021859600001441.

Haghighatafshar, S., la Cour Jansen, Jes, Aspegren, H., Lidström, V., Mattsson, A., & Jönsson, K. (2014). Storm-water management in Malmö and Copenhagen with regard to climate change scenarios. *Vatten*, *70*, 159–168.

Hankin, B., Waller, S., Astle, G., & Kellagher, R. (2008). Mapping space for water: screening for urban flash flooding. *J. Flood Risk Manage.*, *1*, 13–22. doi:10.1111/j.1753-318X.2008.00003.x.

Hénonin, J., Hongtao, M., Zheng-Yu, Y., Hartnack, J., Havnø, K., Gourbesville, P., & Mark, O. (2013). Citywide multi-grid urban flood modelling: The July 2012 flood in Beijing. *Urban Water J.*, *12*, 52–66. doi:10.1080/1573062X.2013.851710.

Horritt, M., & Bates, P. (2001). Effects of spatial resolution on a raster based model of flood flow. *J. Hydrol.*, *253*, 239–249. doi:10.1016/S0022-1694(01)00490-5.

Hunter, N. M., Bates, P. D., Neelz, S., Pender, G., Villanueva, I., Wright, N. G., Liang, D., Falconer, R. A., Lin, B., Waller, S., Crossley, A. J., & Mason, D. C. (2008). Benchmarking 2D hydraulic models for urban flooding. *P. I. Civil Eng.-Wat.*

M., *161*, 13–30. doi:10.1680/wama.2008.161.1.13.

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environ. Modell. Softw.*, *21*, 602–614. doi:10.1016/j.envsoft.2006.01.004.

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., Gericke, A., Neal, J., Aerts, J. C. J. H., & Ward, P. J. (2012). Comparative flood damage model assessment: Towards a European approach. *Nat. Hazards Earth Syst. Sci.*, *12*, 3733–3752. doi:10.5194/nhess-12-3733-2012.

Jordi + Kolb AG (2008). *Schadenkarte Unwetter 2007*. Münsingen, Schweiz.

Kaźmierczak, A., & Cavan, G. (2011). Surface water flooding risk to urban communities: Analysis of vulnerability, hazard and exposure. *Landscape Urban Plan.*, *103*, 185–197. doi:10.1016/j.landurbplan.2011.07.008.

Kipfer, A., Kienholz, C., & Liener, S. (2012). Ein neuer Ansatz zur Modellierung von Oberflächenabfluss. In G. Koboltschnig, J. Hübl, & J. Braun (Eds.), *INTERPRAEVENT 2012 — Conference Proceedings* (pp. 179–189). International Research Society INTERPRAEVENT.

Kipfer, A., Schönthal, E., Liener, S., & Gsteiger, P. (2015). *Oberflächenabflusskarte Kanton Luzern: Bericht*. Bern, Schweiz: geo7 AG.

Kron, W., Steuer, M., Löw, P., & Wirtz, A. (2012). How to deal properly with a natural catastrophe database — analysis of flood losses. *Nat. Hazards Earth Syst. Sci.*, *12*, 535–550. doi:10.5194/nhess-12-535-2012.

Kuhnert, M., Voinov, A., & Seppelt, R. (2005). Comparing raster map comparison algorithms for spatial modeling and analysis. *Photogramm. Eng. Rem. S.*, *71*, 975–984.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K., & Sherstyukov, B. (2014). Flood risk and climate change: global and regional perspectives. *Hydrolog. Sci. J.*, *59*, 1–28. doi:10.1080/02626667.2013.857411.

L'homme, J., Sayers, P., Gouldby, B., Samuels, P., Wills, M., & Mulet-Marti, J. (2008). Recent development and application of a rapid flood spreading method. In P. Samuels, S. Huntington, W. Allsop, & J. Harrop (Eds.), *Flood risk management: research and practice* Proceedings of the European conference on flood risk management research into practice (pp. 15–24). Taylor and Francis Group.

López-Vicente, M., Pérez-Bielsa, C., López-Montero, T., Lambán, L. J., & Navas, A. (2014). Runoff simulation with eight different flow accumulation algorithms: recommendations using a spatially distributed and open-source model. *Environ. Modell. Softw.*, *62*, 11–21. doi:10.1016/j.envsoft.2014.08.025.

Löwe, R., Urich, C., Sto. Domingo, N., Mark, O., Deletic, A., & Arnbjerg-Nielsen, K. (2017). Assessment of urban pluvial flood risk and efficiency of adaptation options through simulations — A new generation of urban planning tools. *J. Hydrol.*, *550*, 355–367. doi:10.1016/j.jhydrol.2017.05.009.

LUBW (2016). *Leitfaden kommunales Starkregenrisikomanagement in Baden-Württemberg*. Karlsruhe, Deutschland: Landesanstalt für Umwelt, Messungen und Naturschut Baden-Württemberg (LUBW). URL: http://www.lubw.baden-wuerttemberg.de/servlet/is/261161.

Maksivomić, Č., Prodanović, D., Boonya-Aroonnet, S., Leitão, J. P., Djordjević, S., & Allitt, R. (2009). Overland flow and pathway analysis for modelling of urban pluvial flooding. *J. Hydraul. Res.*, *47*, 512–523. doi:10.3826/jhr.2009.3361.

McCuen, R. H. (2016). *Hydrologic analysis and design*. (4th ed.). Hoboken, USA: Pearson Higher Education.

Mein, R. G., & Larson, C. L. (1973). Modeling infiltration during a steady rain. *Water Resour. Res.*, *9*, 384–394. doi:10.1029/WR009i002p00384.

Meon, G., Stein, K., Förster, K., & Riedel, G. (2009). *Untersuchung starkregengefährdeter Gebiete: Abschlussbericht zum Forschungsprojekt*. Braunschweig, Deutschland: Technische Universität Braunschweig and Leichtweiss-Institut für Wasserbau.

Merwade, V., Olivera, F., Arabi, M., & Edleman, S. (2008). Uncertainty in flood inundation mapping: current Issues and future directions. *J. Hydrol. Eng.*, *13*, 608–620. doi:10.1061/(ASCE)1084-0699(2008)13:7(608).

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article — "Assessment of economic flood damage". *Nat. Hazards Earth Syst. Sci.*, *10*, 1697–1724. doi:10.5194/nhess-10-1697-2010.

MeteoSwiss (2014). *Räumliche Daten CombiPrecip*. Swiss Federal Office of Meteorology and Climatology. URL: http://www.meteoswiss.admin.ch/home/services-and-publications/produkte.subpage.html/en/data/products/2014/raeumliche-daten-combiprecip.html (last accessed 03.07.2017).

Mitasova, H., Thaxton, C., Hofierka, J., McLaughlin, R., Moore, A., & Mitas, L. (2004). Path sampling method for modeling overland water flow, sediment transport, and short term terrain evolution in Open Source GIS. In C. T. Miller, M. W. Farthing, G. F. Pinder, & W. G. Gray (Eds.), *Proceedings of the XVth International Conference on Computational Methods in Water Resources (CMWR XV)* (pp. 1479–1490). Chapel Hill, USA: Elsevier. doi:10.1016/S0167-5648(04)80159-X.

Neal, J., Villanueva, I., Wright, N., Willis, T., Fewtrell, T., & Bates, P. (2012). How much physical complexity is needed to model flood inundation? *Hydrol. Process.*, *26*, 2264–2282. doi:10.1002/hyp.8339.

Néelz, S., & Pender, G. (2013). *Benchmarking the latest generation of 2D hydraulic modelling packages: Report – SC120002*. Bristol, UK: Environment Agency.

Neteler, M., Bowman, M. H., Landa, M., & Metz, M. (2012). GRASS GIS: A multi-purpose open source GIS. *Environ. Modell. Softw.*, *31*, 124–130. doi:10.1016/j.envsoft.2011.11.014.

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., & Saleska, S. (2011). Height above the nearest drainage — a hydrologically relevant new terrain model. *J. Hydrol.*, *404*, 13–29. doi:10.1016/j.jhydrol.2011.03.051.

O'Brian, J. S. (2009). *FLO-2D Reference Manual: Version 2009*. Nutrioso, USA: FLO-2D Software Inc. URL: https://www.flo-2d.com/download/.

van Ootegem, L., Verhofstadt, E., van Herck, K., & Creten, T. (2015). Multivariate pluvial flood damage models. *Environ. Impact Asses.*, *54*, 91–100. doi:10.1016/j.eiar.2015.05.005.

Panziera, L., Gabella, M., Zanini, S., Hering, A., Germann, U., & Berne, A. (2016). A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland. *Hydrol. Earth Syst. Sc.*, *20*, 2317–2332. doi:10.5194/hess-20-2317-2016.

Pappenberger, F., Frodsham, K., Beven, K. J., Romanowicz, R., & Matgen, P. (2007). Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. *Hydrol. Earth Syst. Sc.*, *11*, 739–752. doi:10.5194/hess-11-739-2007.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ. Modell. Softw.*, *79*, 214–232. doi:10.1016/j.envsoft.2016.02.008.

Pilesjö, P., & Hasan, A. (2014). A triangular form-based multiple flow algorithm to estimate overland flow distribution and accumulation on a digital elevation model. *T. GIS*, *18*, 108–124. doi:10.1111/tgis.12015.

Pitt, M. (2008). *The Pitt Review: Learning lessons from the 2007 floods: An independent review by Sir Michael Pitt*. London, UK: Cabinet Office.

Planchon, O., & Darboux, F. (2002). A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, *46*, 159–176. doi:10.1016/S0341-8162(01)00164-3.

Quinn, P., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol. Process.*, *5*, 59–79.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org.

Rawls, W. J., Brakensiek, D. L., & Miller, N. (1983). Green–Ampt infiltration parameters from soils data. *J. Hydraul. Eng.-ASCE*, *109*, 62–70. doi:10.1061/(ASCE)0733-9429(1983)109:1(62).

Röthlisberger, V., Zischg, A. P., & Keiler, M. (2017). Identifying spatial clusters of flood exposure to support decision making in risk management. *Sci. Total Environ.*, *598*, 593–603. doi:10.1016/j.scitotenv.2017.03.216.

Rözer, V., Müller, M., Bubeck, P., Kienzler, S., Thieken, A., Pech, I., Schröter, K., Buchholz, O., & Kreibich, H. (2016). Coping with pluvial floods by private households. *Water*, *8*, 304. doi:10.3390/w8070304.

Rüttimann, D., & Egli, T. (2010). *Wegleitung punktuelle Gefahrenabklärung Oberflächenwasser*. St. Gallen, Schweiz: Naturgefahrenkommission des Kantons St. Gallen.

Sampson, C. C., Bates, P. D., Neal, J. C., & Horritt, M. S. (2013). An automated routing methodology to enable direct rainfall in high resolution shallow water models. *Hydrol. Process.*, *27*, 467–476. doi:10.1002/hyp.9515.

Sampson, C. C., Fewtrell, T. J., Duncan, A., Shaad, K., Horritt, M. S., & Bates, P. D. (2012). Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. *Adv. Water Resour.*, *41*, 1–17. doi:10.1016/j.advwatres.2012.02.010.

Savage, J. T. S., Bates, P., Freer, J., Neal, J., & Aronica, G. (2016a). When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrol. Process.*, *30*, 2014–2032. doi:10.1002/hyp.10749.

Savage, J. T. S., Pianosi, F., Bates, P. D., Freer, J., & Wagener, T. (2016b). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resour. Res.*, *52*, 9146–9163. doi:10.1002/2015WR018198.

Schmocker-Fackel, P., Naef, F., & Scherrer, S. (2007). Identifying runoff processes on the plot and catchment scale. *Hydrol. Earth Syst. Sc.*, *11*, 891–906. doi:10.5194/hess-11-891-2007.

Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., & Pappenberger, F. (2009). Progress in integration of remote sensing–derived flood extent and stage data and hydraulic models. *Rev. Geophys.*, *47*, RG4001. doi:10.1029/2008RG000274.

Seibert, J., & McGlynn, B. L. (2007). A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resour. Res.*, *43*. doi:10.1029/2006WR005128.

Sideris, I. V., Gabella, M., Erdin, R., & Germann, U. (2014). Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Q.J.R. Meteorol. Soc.*, *140*, 1097–1111. doi:10.1002/qj.2188.

Spekkers, M., Rözer, V., Thieken, A., ten Veldhuis, M.-c., & Kreibich, H. (2017). A comparative survey of the impacts of extreme rainfall in two international case studies. *Nat. Hazards Earth Syst. Sci.*, *17*, 1337–1355. doi:10.5194/nhess-17-1337-2017.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Nat. Hazards Earth Syst. Sci.*, *14*, 2531–2547. doi:10.5194/nhess-14-2531-2014.

Steinbrich, A., Leistert, H., & Weiler, M. (2016). Model-based quantification of runoff generation processes at high spatial and temporal resolution. *Environ. Earth Sci.*, *75*, 1423. doi:10.1007/s12665-016-6234-9.

Stephens, E., Schumann, G., & Bates, P. (2014). Problems with binary pattern measures for flood model evaluation. *Hydrol. Process.*, *28*, 4928–4937. doi:10.1002/hyp.9979.

swisstopo (2017a). *swissALTI3D: The high precision digital elevation model of Switzerland*. Swiss Federal Office of Topography. URL: https://shop.swisstopo.admin.ch/en/products/height_models/alti3D (last accessed 03.07.2017).

swisstopo (2017b). *SWISSIMAGE: The digital color orthophotomosaic of Switzerland*. Swiss Federal Office of Topography. URL: https://shop.swisstopo.admin.ch/en/products/images/ortho_images/SWISSIMAGE (last accessed 03.07.2017).

Teng, J., Jakeman, A. J., Vaze, J., Croke, B., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environ. Modell. Softw.*, *90*, 201–216. doi:10.1016/j.envsoft.2017.01.006.

Thieken, A. H., Kreibich, H., Müller, M., & Merz, B. (2007). Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002. *Hydrolog. Sci. J.*, *52*, 1016–1037. doi:10.1623/hysj.52.5.1016.

Tyrna, B., Assmann, A., Fritsch, K., & Johann, G. (2017). Large-scale high-resolution pluvial flood hazard mapping using the raster-based hydrodynamic two-dimensional model FloodAreaHPC. *J. Flood Risk Manage.*, *42*, 19. doi:10.1111/jfr3.12287.

Tyrna, B. G., & Hochschild, V. (2010). Modellierung von lokalen Überschwemmungen nach Starkniederschlägen. In J. Strobl, T. Blaschke, & G. Griesebner (Eds.), *Angewandte Geoinformatik 2010* (pp. 325–334).

Ward, R. C., & Robinson, M. (2000). *Principles of hydrology*. (4th ed.). London, UK: McGraw-Hill.

Wechsler, S. P. (2007). Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrol. Earth Syst. Sc.*, *11*, 1481–1500. doi:10.5194/hess-11-1481-2007.

Yu, D., & Coulthard, T. J. (2015). Evaluating the importance of catchment hydrological parameters for urban surface water flood modelling using a simple hydro-inundation model. *J. Hydrol.*, *524*, 385–400. doi:10.1016/j.jhydrol.2015.02.040.

Zhou, Q., & Liu, X. (2002). Error assessment of grid-based flow routing algorithms used in hydrological models. *Int. J. Geogr. Inf. Sci.*, *16*, 819–842. doi:10.1080/13658810210149425.

Zhou, Q., Mikkelsen, P. S., Halsnæs, K., & Arnbjerg-Nielsen, K. (2012). Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. *J. Hydrol.*, *414–415*, 539–549. doi:10.1016/j.jhydrol.2011.11.031.

Zischg, A. P., Mosimann, M., Bernet, D. B., & Röthlisberger, V. (2018). Validation of 2D flood models with insurance claims. *J. Hydrol.*, *557*, 350–361. doi:10.1016/j.jhydrol.2017.12.042.

- Surface water floods in rural areas are investigated
- Value of single deterministic simulations with uncalibrated models is assessed
- Four variously complex raster-based flood inundation models are employed
- Models are tested on four artificial surfaces
- Inundated areas are predicted for eight real-world case studies with each model